Sylvia Wassertheil-Smoller
Jordan Smoller

Biostatistică și Epidemiologie

Un primer pentru sănătate și profesioniști biomedicali

Ediția a patra

Springer

Biostatistică și Epidemiologie

Ediția a patra

Sylvia Wassertheil-Smoller

Departamentul de Epidemiologie, Colegiul de Medicină Albert Einstein, Bronx, NY, SUA

Jordan Smoller

Departamentul de Psihiatrie și Centrul de Cercetare Genetică Umană, Massachusetts General Hospital, Boston, MA, SUA

Biostatistica si Epidemiologie

Un primer pentru sănătate și Profesioniști în domeniul biomedical

Ediția a patra

Cu 37 de ilustrații

Springer

Sylvia Wassertheil-Smoller
Department of Epidemiology
Albert Einstein College of Medicine
Bronx, NY, USA

Jordan Smoller
Department of Psychiatry and Center
for Human Genetic Research
Massachusetts General Hospital
Boston, MA, USA

ISBN 978-1-4939-2133-1 ISBN 978-1-4939-2134-8 (eBook)
DOI 10.1007/978-1-4939-2134-8

Număr de control al Bibliotecii Congresului: 2014952775

Springer New York Heidelberg Dordrecht Londra
© Springer Science+Business Media New York 1990, 1995, 2004, 2015
Această lucrare este supusă dreptului de autor. Toate drepturile sunt rezervate de către Editor, indiferent dacă este vorba de materialul integral sau parțial, în special drepturile de traducere, retipărire, reutilizare a ilustrațiilor, recitare, difuzare, reproducere pe microfilme sau în orice alt mod fizic și transmitere sau stocare a informațiilor. și regăsire, adaptare electronică, software de calculator sau prin metodologie similară sau diferită cunoscută acum sau dezvoltată în continuare. Sunt exceptate de la această rezervă legală fragmentele scurte în legătură cu recenzii sau analize academice sau materiale furnizate special în scopul de a fi introduse și executate pe un sistem informatic, pentru uzul exclusiv de către cumpărătorul lucrării. Duplicarea acestei publicații sau a unor părți ale acesteia este permisă numai în conformitate cu prevederile Legii privind drepturile de autor a locației Editorului, în versiunea sa actuală, iar permisiunea de utilizare trebuie obținută întotdeauna de la Springer. Permisiunile de utilizare pot fi obținute prin RightsLink la Centrul de autorizare a drepturilor de autor. Încălcările sunt pasibile de urmărire penală în conformitate cu Legea drepturilor de autor respectivă.
Utilizarea numelor descriptive generale, a numelor înregistrate, a mărcilor comerciale, a mărcilor de servicii etc. în această publicație nu implică, chiar și în absența unei declarații specifice, că astfel de nume sunt scutite de legile și reglementările de protecție relevante și, prin urmare, sunt gratuite pentru general. utilizare.
Deși sfaturile și informațiile din această carte sunt considerate a fi adevărate și exacte la data publicării, nici autorii, nici editorii, nici editorul nu pot accepta nicio responsabilitate legală pentru eventualele erori sau omisiuni care pot fi făcute. Editorul nu oferă nicio garanție, expresă sau implicită, cu privire la materialul conținut aici.

Imprimat pe hârtie fără acid

Springer face parte din Springer Science+Business Media (www.springer.com)

Pentru Alexis și Ava

PREFAȚĂ LA EDIȚIA A PATRA

Această carte, prin numeroasele sale ediții, a continuat să se adapteze la domeniile în evoluție de cercetare în epidemiologie și statistică, menținând în același timp obiectivul inițial de a fi neamenințătoare, de înțeles și accesibilă celor cu experiență limitată sau nu în matematică. Noi domenii sunt acoperite în cea de-a patra ediție și includ un nou capitol despre predicția riscului, reclasificarea riscului și evaluarea biomarkerilor, material nou privind analizele de propensiune și un capitol extins și actualizat cu privire la epidemiologia genetică.

Odată cu secvențierea genomului uman, a existat o înflorire a cercetării asupra bazei genetice a sănătății și bolii, și în special a interacțiunilor dintre gene și expunerile la mediu. Literatura medicală în epidemiologia genetică este în expansiune enormă și este necesară o anumită cunoaștere a modelelor epidemiologice și o cunoaștere a metodelor statistice utilizate în astfel de cercetări pentru a putea aprecia noile descoperiri. Astfel, această ediție include un nou capitol despre epidemiologia genetică. Un astfel de material nu se găsește de obicei în cărțile de epidemiologie sau statistică de prim nivel, dar este prezentat aici într-un mod de bază și, sperăm, ușor de înțeles , pentru cei nefamiliarizați sau puțin familiarizați cu domeniul.

Un alt capitol nou se referă la predicția riscului, care este importantă atât din perspectivă clinică individuală pentru evaluarea riscului pacientului în vederea instituirii măsurilor preventive adecvate, cât și din perspectivă de sănătate publică pentru a evalua nevoile unei populații. Pe măsură ce obținem o mai bună înțelegere a biologiei implicate în procesele bolii, noi biomarkeri ai bolii sunt investigați fie pentru a prezice boala, fie pentru a servi drept ținte pentru noi măsuri terapeutice. Este important să se evalueze astfel de biomarkeri pentru a vedea dacă îmbunătățesc de fapt predicția riscului dincolo de cea obținută din factorii de risc tradiționali. Noul capitol explică logica și tehnicile statistice utilizate pentru astfel de evaluări.

Studiul clinic randomizat este „standardul de aur” pentru dovezi privind cauzalitatea și compararea tratamentelor. Cu toate acestea, nu putem face studii clinice în toate domeniile, fie din cauza problemelor de fezabilitate, a costurilor ridicate sau a dimensiunii eșantionului și a duratei de urmărire necesare pentru a trage concluzii valide. Astfel, trebuie să ne bazăm adesea pe dovezile din studii

observaționale care pot
fi supusă confuziei. Analiza înclinației este o tehnică analitică din ce în ce mai utilizată pentru a controla confuzia, iar ediția a 4-a oferă o explicație cuprinzătoare a metodelor implicate.

A fost adăugat și material nou la mai multe capitole existente.

Obiectivele principale ale edițiilor anterioare încă se aplică. Prezentarea - materialului este menită să ofere o înțelegere a principiilor de bază, precum și linii directoare practice despre „cum să o faci" și „cum să o interpretezi". Subiectele incluse sunt cele care sunt cele mai frecvent utilizate sau la care se face referire în literatură. Există câteva caracteristici de reținut care pot ajuta cititorul în utilizarea acestei cărți:

(a) Cartea începe cu o discuție despre filozofia și logica științei și principiile care stau la baza testării a ceea ce credem față de realitatea experiențelor noastre. Deși o astfel de discuție, în sine, nu va ajuta cititorul să „facă un test t ", credem că este important să oferim o introducere în cadrul de bază al domeniului epidemiologiei și statisticii, pentru a înțelege de ce facem ceea ce facem.

(b) Multe dintre subsecțiuni sunt singure; adică cititorul poate apela la subiectul care îl interesează și poate citi materialul în ordine secvențială. Astfel, cartea poate fi folosită de cei care au nevoie de ea în scopuri speciale. Cititorul este liber să sări peste acele subiecte care nu sunt de interes fără a fi prea mult împiedicat în continuarea lecturii. Ca urmare, există o oarecare redundanță. Totuși, în experiența noastră de predare, am constatat că este mai bine să greșim din partea redundanței decât din partea rarității.

(c) Referințele încrucișate la alte secțiuni relevante sunt incluse atunci când sunt necesare explicații suplimentare. Atunci când dezvoltarea unui subiect depășește sfera acestui text, cititorul este îndrumat către alte cărți care tratează materialul în profunzime sau la un nivel matematic superior. O listă de texte recomandate este oferită aproape de sfârșitul cărții.

(d) Anexele oferă exemple de calcule pentru diverse statistici descrise în text. Acest lucru face o citire mai ușoară a textului, oferind în același timp cititorului instrucțiuni mai specifice despre cum să facă de fapt unele dintre calcule.

Ediţiile anterioare au crescut din feedback-ul studenţilor care au indicat că apreciază claritatea şi concentrarea pe subiecte legate în mod specific de munca lor. Cu toate acestea, unii utilizatori au ratat acoperirea mai multor subiecte importante. În consecinţă, au fost adăugate secţiuni pentru a include un capitol complet despre măsurile calităţii vieţii şi diverse scale psihologice, care sunt din ce în ce mai utilizate în studiile clinice; o extindere a capitolului despre probabilitate, cu introducerea mai multor metode neparametrice; clarificarea unor concepte care au fost abordate mai concis anterior ; şi adăugarea mai multor anexe (care oferă eşantion de calcule ale testului exact al lui Fisher , testului Kruskal-Wallis şi diverşi indici de fiabilitate şi de receptivitate a scalelor utilizate în măsurile calităţii vieţii).

Este nevoie de un echilibru delicat pentru a menţine cartea concisă şi de bază şi totuşi pentru a o face suficient de incluzivă pentru a fi utilă unui public larg. Sperăm că această carte va fi utilă diverselor grupuri de oameni din domeniul sănătăţii, precum şi celor din domenii conexe. Materialul este destinat: (1) medicilor care efectuează cercetări clinice, precum şi celor care fac cercetări de bază; (2) studenţi — medicină, facultate şi absolvenţi; (3) personal de cercetare în diferite capacităţi; (4) cei interesaţi de domeniul în creştere al epidemiologiei genetice şi care doresc să poată citi cercetarea genetică sau care doresc să colaboreze în cercetarea genetică; şi (5) oricine este interesat de logica şi metodologia biostatisticii, epidemiologiei şi epidemiologiei genetice. Principiile şi metodele descrise aici sunt aplicabile în diferite domenii de fond, inclusiv medicină, sănătate publică, psihologie şi educaţie. Desigur, nu toate subiectele care sunt relevante în mod specific pentru fiecare dintre aceste discipline pot fi acoperite în acest scurt text.

Bronx, NY, USA — Sylvia Wassertheil-Smoller

Boston, MA, USA — Jordan W. Smoller

MULȚUMIRI

Vreau să-mi exprim recunoștința pentru predarea inspirată a doctorului Jacob Cohen, acum decedat, care m-a pornit pe acest drum și colegilor și studenților mei de la Colegiul de Medicină Albert Einstein, care o fac amuzantă.

Aprecierea mea se adresează acelor colegi care au criticat edițiile anterioare și mulțumiri speciale sunt adresate Dr. Aileen McGinn, Dr. Gloria Ho, Dr. Tao Wang și Dr. Kenny Ye pentru ajutorul acordat în editarea noului material din această ediție și pentru întotdeauna sugestii înțelepte.

Din păcate, regretatul meu soț, Walter Austerer, nu este aici pentru a se bucura de această nouă ediție, dar doresc să-i onorez memoria și răbdarea, dragostea și sprijinul pe care le-a oferit fără greș prin edițiile anterioare.

În cele din urmă, vreau să spun ce mare privilegiu și plăcere este să lucrez cu coautorul meu, care este fiul meu, Jordan Smoller. Viața rareori aduce astfel de recompense.

Bronx, New York Sylvia Wassertheil-Smoller

CUPRINS

PREFAȚĂ LA EDIȚIA A PATRA vii
MULȚUMIRI xi
BIOGRAFIA AUTORULUI xix

CHAPTER 1. METODA ȘTIINȚIFICĂ 1
1.1 Logica raționamentului științific 1
1.2 Variabilitatea fenomenelor necesită o analiză statistică 6
1.3 Inferență inductivă: Statistica ca tehnologie a metodei științifice 7
1.4 Proiectarea studiilor 8
1.5 Cum se cuantifică variabilele 10
1.6 Ipoteza nulă 10
1.7 De ce testăm ipoteza nulă? 11
1.8 Tipuri de erori 13
1.9 Nivel de semnificație și tipuri de eroare 15
1.10 Consecințele erorilor de tip I și de tip II 15

CHAPTER 2. PUȚINĂ DE PROBABILITATE 17
2.1 Ce este probabilitatea? 17
2.2 Combinarea probabilităților 18
2.3 Probabilitate condiționată 20
2.4 Probabilitatea Bayesiană 21
2.5 Cote și probabilitate 22
2.6 Raportul de probabilitate 23
2.7 Rezumatul probabilității 24

CHAPTER 3. CELE MAI MULT DESPRE STATISTICI 27
3.1 Chi-pătrat pentru 2 × 2 mese 27
3.2 Testul McNemar 31
3.3 Kappa 33
3.4 Descrierea unei populații: Utilizarea standardului Abaterea 34

3.5 Semnificația abaterii standard: distribuția normală 37
3.6 Diferența dintre abaterea standard și eroarea standard 39
3.7 Eroare standard a diferenței dintre două mijloace 42
3.8 Scorurile Z și distribuția normală standardizată 44
3.9 Statistica t 47
3.10 Valori eșantioane și valori ale populației revizuite 48
3.11 O întrebare de încredere 49
3.12 Limite de încredere și intervale de încredere 51
3.13 Grade de libertate 52
3.14 Intervale de încredere pentru proporții 52
3.15 Intervale de încredere în jurul diferenței dintre două mijloace 53
3.16 Comparații între două grupuri 54
3.17 Testul Z pentru compararea a două proporții 55
3.18 t-Test pentru diferența dintre mijloacele a două grupuri independente: principii 57
3.19 Cum se face un t -test: un exemplu 58
3.20 Testul t de perechi potrivite 60
3.21 Când să nu faceți o mulțime de teste t: problema testelor multiple de semnificație 61
3.22 Analiza variației: comparație între mai multe grupuri 63
3.23 Principii care stau la baza analizei variației 63
3.24 Procedura Bonferroni: O abordare pentru a face comparații multiple 66
3.25 Analiza variației atunci când există două variabile independente: ANOVA cu doi factori 68
3.26 Interacțiunea între două variabile independente 69
3.27 Exemplu de ANOVA bidirecțional 70
3.28 Testul Kruskal–Wallis pentru a compara mai multe grupuri 71
3.29 Asociere și cauza: coeficientul de corelație 71
3.30 Câteva puncte de reținut despre corelație 74
3.31 Căi cauzale 75

3.32
3.33 Regresia 76
Legătura dintre regresia liniară
și coeficientul de corelație 79
3.34 Regresia liniară multiplă 79
3.35 Rezumat până acum 81

CAPITOLUL 4. CELE MAI MULT DESPRE EPIDEMIOLOGIE 83
4.1 Utilizările epidemiologiei 83
4.2 Câteva concepte epidemiologice: ratele mortalității 84
4.3 Tarife ajustate în funcție de vârstă 86
4.4 Incidență și prevalență 88
4.5 Rata de mortalitate standardizată 90
4.6 Persoană-ani de observație 90
4.7 Variabile dependente și independente 92
4.8 Tipuri de studii 92
4.9 Analize transversale versus longitudinale ale datelor 93
4.10 Măsuri ale riscului relativ: concluzii din perspectiva
Studii (Studiul Framingham) 97

4.11 Calculul riscului relativ din studii prospective 99
4.12 Odds Ratio: Estimarea riscului relativ din
Studii de caz-control 100
4.13
4.14 Risc atribuibil 103
4.15 Bias de răspuns 105
4.16 Variabile de confuzie 108
4.17 Potrivire 108
4.18 Regresie logistică multiplă 110
4.19 Analiza supraviețuirii: metodele tabelului de viață 112
4.20 115 de pericole proporționale Cox
Intervale de încredere și statistică suprapuse
Semnificația 117
4.21 Confuzie prin indicația 117
4.22 Analiza înclinației 118
4.23 Selectarea variabilelor pentru modelele multivariate 123
4.24 Interacțiuni: modele aditive și multiplicative 125
4.25 Relații neliniare: formă J sau formă U 129

CHAPTER 5. MAI MULT DESPRE PROIECTARE 133
5.1 Sensibilitate, specificitate și concepte înrudite 133
5.2 Cutoff Point și efectele sale asupra sensibilității și specificității 140

CHAPTER 6. CELE MAI MULT DESPRE TRESTILE CLINICE 143
6.1 Caracteristicile studiilor clinice randomizate 143
6.2 Scopurile randomizării 145
6.3 Cum se efectuează o atribuire aleatorie 146
6.4 Test cu două cozi versus test cu o singură coadă 147
6.5 Studiul clinic ca „Standard de aur” 148
6.6 Regresia spre medie 149
6.7 Analiza intentiei de tratare 152
6.8 Cât de mare ar trebui să fie studiul clinic? 153
6.9 Ce este implicat în calculul dimensiunii eșantionului? 155
6.10 Cum se calculează dimensiunea eșantionului pentru diferența dintre două proporții 159
6.11 Cum se calculează dimensiunea eșantionului pentru a testa diferența dintre două mijloace 160
163
164

CHAPTER 7. CELE MAI MULT DESPRE CALITATEA VIEI 164
7.1 Construcție la scară 166
7.2 Fiabilitate 167
7.3 Valabilitate 169
7.4 Receptivitatea 171
7.5 Câteva capcane potențiale 171
173

CHAPTER 8. MAI MULT DESPRE EPIDEMIOLOGIA GENETICĂ 175
8.1 O nouă eră științifică 176
8.2 Prezentare generală a epidemiologiei genetice 179
8.3 Studii pe gemeni 180
8.4 Studii de legături și asocieri 182
8.5 Scor LOD: Statistică de legătură 183
8.6 Studii de Asociere
8.7 Studii ale asociației de gene candidate
8.8 Stratificarea populației sau Structura populației

8.9 Asociația bazată pe familie și testul de dezechilibru al transmisiei (TDT) 186
8.10 Studii de asociere la nivel de genom: GWAS 189
8.11 Controlul calității GWAS și echilibrul Hardy–Weinberg 191
8.12 Diagrame cuantile după cuantile sau diagrame QQ 192
8.13 Problema falselor pozitive 195
8.14 Problema falselor negative 197
8.15 Manhattan Plots 198
8.16 Scoruri Polygene 199
8.17 Variante rare și secvențierea genomului 200
8.18 Analiza studiilor de variante rare 202
8.19 Ce este într-un nume? SNP-uri și gene 203

CHAPTER 9. PREDICȚIA RISCURILOR ȘI CLASIFICAREA RISCURILOR 205
9.1 Predicția riscului 205
9.2 Valoarea aditivă a unui biomarker: calcul a riscului prezis 207
9.3 Indicele net de îmbunătățire a reclasificării 212
9.4 NRI fără categorie 213
9.5 Îmbunătățirea integrată a discriminării (IDI) 214
9.6 C-Statistic 215
9.7 Avertismente 216
9.8 Rezumatul 216

CAPITOLUL 10. ETICA ŞI STATISTICA CERCETĂRII 217
10.1 Ce legătură are statisticile cu ea? 217
10.2 Protecția subiectelor de cercetare umane 218
10.3 Consimțământ informat 220
10.4 Echilibrul 222
10.5 Integritatea cercetării 222
10.6 Politici de autor 223
10.7 Comisiile de monitorizare a datelor și a siguranței 224
10.8 Rezumatul 224

Postscript CÂTEVA COMENTARII DE DESPĂZIRE LA IMPACTUL EPIDEMIOLOGIEI ASUPRA OMULUI VIEȚI 225

Anexa 1. VALORI CRITICE ALE CHI-PĂTRATULUI, Z și t 227

Anexa 2. TESTUL EXACT LUI FISHER 229

Appendix 3. KRUSKAL–WALLIS NEPARAMETRIC TEST PENTRU COMPARAREA MAI MULTE GRUPE 231

Appendix 4. CUM SE CALCULEAZĂ O CORELATIE COEFICIENTUL 233

Appendix 5. AJUSTARE Vârsta 235

Appendix 6. DETERMINAREA ADECVĂRII DE SCHIMBAREA SCORURILOR 239

REFERINȚE 243
LECTURI SUGERATE 249
INDEX 253

BIOGRAFIA AUTORULUI

Sylvia Wassertheil-Smoller, Ph.D. este profesor de Epidemiologie la Departamentul de Epidemiologie și Sănătatea Populației la Albert Einstein College of Medicine și este Dorothy and Wil liam Manealoff Foundation și Molly Rosen Chair in Social Medicine, Emerita. Pe lângă predarea ei, domeniile ei de cercetare acoperă atât bolile cardiovasculare, cât și epidemiologia cancerului. Ea a fost un lider în pământ-marca studiile clinice în prevenirea bolilor de inimă și a accidentului vascular cerebral și în studii prospective de colaborare majore naționale și internaționale. Ea a câștigat premii pentru mentorat studenți și profesori juniori, precum și Premiul Einstein Spirit of Achievement. Ea locuiește în New York.

Jordan Smoller, MD, Sc.D. este profesor de psihiatrie la Harvard Medical School și profesor la Departamentul de Epidemiologie la Harvard School of Public Health din Boston. El este director al Unității de Genetică Psihiatrică și Neurodezvoltare din Centrul de Cercetare a Geneticii Umane a Spitalului General din Massachusetts (MGH). El este, de asemenea, membru asociat al Institutului Broad și om de știință principal la Centrul Stanley pentru Cercetare Psihiatrică al Broad . Obiectivul cercetării Dr. Smoller a fost identificarea determinanților genetici ai bolilor psihiatrice ale copilăriei și adulților.

ordinelor și a factorilor genetici comuni afecţiunilor psihiatrice multiple. El este beneficiarul a numeroase premii de cercetare; un autor a peste 200 de articole științifice, capitole de carte și recenzii; precum și autorul cărții The Other Side of Normal (HarperCollins/William Morrow, 2012). Locuiește cu familia în Boston.

Chapter 1
METODA ŞTIINŢIFICĂ

Știința este construită cu fapte, așa cum o casă este cu pietre. Dar o colecție de fapte nu este mai mult o știință decât un morman de pietre este o casă.

Jules Henri Poincare
La Science et l ' Hypothese (1908)

1.1 Logica raționamentului științific

Scopul științei este de a descoperi „adevărul". Cum să decidem că ceva este adevărat? Avem două instrumente la dispoziție pentru a continua cercetarea științifică:

Avem simțurile noastre, prin care experimentăm lumea și facem observații .

Avem capacitatea de a raționa, ceea ce ne permite să facem inferențe logice .

În știință impunem logica acestor observații.

În mod clar, avem nevoie de ambele instrumente. Toată logica din lume nu va crea o observație, iar toate observațiile individuale din lume nu vor crea în sine o teorie. Există două tipuri de relații între mintea științifică și lume și două tipuri de logică pe care le impunem – deductivă și inductivă – așa cum este ilustrat în Figura 1.1 .

În inferența deductivă , deținem o teorie și, pe baza ei, facem o predicție a consecințelor acesteia. Adică, prezicem care ar trebui să fie observațiile. De exemplu, putem susține o teorie a învățării care spune că întărirea pozitivă are ca rezultat o învățare mai bună decât pedeapsa, adică recompensele funcționează mai bine decât pedepsele. Din această teorie, prezicem că studenții la matematică care sunt lăudați pentru răspunsurile corecte în timpul anului se vor descurca mai bine

la examenul final decât cei care sunt pedepsiți pentru

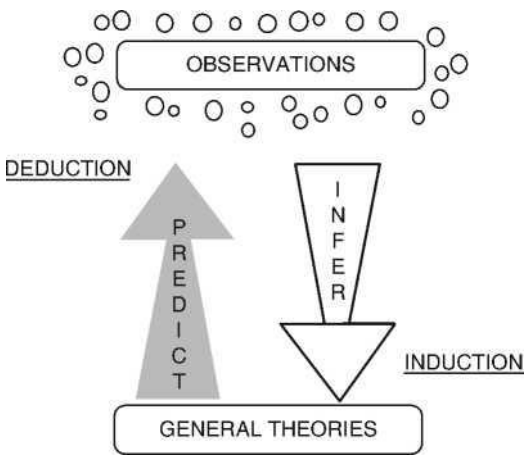


Figura 1.1 Inferența deductivă și inductivă

răspunsurile lor greșite. Trecem de la general, la teorie, la specific, la observații. Aceasta este cunoscută drept metoda ipotetico-deductivă.

În inferența inductivă, trecem de la specific la general. Facem multe observații, discernem un model, facem o generalizare și deducem o explicație. De exemplu, s-a observat la Spitalul General din Viena în anii 1840 că femeile care năsteau mureau cu o rată ridicată de febră puerperală, o generalizare care a provocat teroare la viitoarele mame. Un tânăr medic pe nume Ignaz Phillip Semmelweis a făcut legătura între observația că studenții la medicină care efectuează examinări vaginale au făcut acest lucru imediat după ce au venit din sala de disecție, spălându-se rar pe mâini între ele, cu observația că un coleg care și-a tăiat accidental degetul în timp ce diseca un cadavrul a murit de o boală exact ca cea care ucide mamele. El a dedus explicația că cauza morții a fost introducerea de material cadaveru într-o rană. Consecința practică a acelui salt creativ al imaginației a fost eliminarea febrei puerperale ca flagel al nașterii prin cerea ca medicii să se spele pe mâini înainte de a face o naștere! Abilitatea de a face astfel de salturi creative din generalizări este produsul minților științifice creative.

În general, s-a considerat că epidemiologii folosesc inferența inductivă . De exemplu, cu câteva decenii în urmă, s-a observat că femeile păreau să facă atacuri de cord cu aproximativ 10 ani mai târziu decât bărbații. Un salt creativ al

imaginației a dus la concluzia că hormonii femeilor au fost cei care le-au protejat până la menopauză. EUREKA! Ei au dedus că, dacă estrogenul este bun pentru femei, trebuie să fie bun pentru bărbați și au prezis că observațiile vor corobora această deducere. A fost efectuat un studiu clinic care a dat bărbaților cu risc crescut de atac de cord estrogen în doze destul de mari, 2,5 mg pe zi sau aproximativ de patru ori doza utilizată în prezent la femeile aflate în postmenopauză. Deloc surprinzător, bărbații nu au apreciat efectele secundare, dar în mod surprinzător pentru anchetatori, bărbații din grupul cu estrogeni au avut rate mai mari de boli coronariene și mortalitate decât cei tratați cu placebo. [2] Ceea ce a fost bun pentru gâscă s-ar putea să nu fie atât de bun pentru gâscă. Studiul a fost întrerupt, iar estrogenul ca măsură preventivă a fost abandonat timp de câteva decenii.

În acest timp, multe studii observaționale prospective au indicat că înlocuirea estrogenului administrată femeilor aflate în postmenopauză a redus riscul de boli de inimă cu 30-50%. Aceste observații au condus la concluzia inductivă că înlocuirea hormonală în postmenopauză este protectoare, adică observațiile au condus la teorie. Cu toate acestea, această teorie trebuie testată în studii clinice. Primul astfel de studiu de substituție hormonală la femeile care aveau deja boli de inimă, Studiul de înlocuire a inimii și a estrogenului/progestinului (HERS), nu a găsit nicio diferență în ratele bolilor de inimă între grupul de tratament activ și grupul placebo, dar a găsit o creștere precoce. în evenimentele de boli de inimă în primul an de studiu și un beneficiu ulterior al hormonilor după aproximativ 2 ani. Deoarece acesta a fost un studiu pe femei cu boală cardiacă stabilită, a fost un studiu de prevenire secundară și nu răspunde la întrebarea dacă femeile fără boli de inimă cunoscute ar beneficia de înlocuirea hormonală pe termen lung. Această întrebare a fost abordată de Inițiativa pentru sănătatea femeilor (WHI), care este descrisă într-o secțiune ulterioară.

Scopul exemplului este de a ilustra modul în care observațiile (că femeile suferă de boli de inimă mai târziu decât bărbații) duc la o teorie (că hormonii sunt protectori), care prezice noi observații (că vor exista mai puține atacuri de cord și decese în rândul celor cu hormoni), care poate întări teoria, până când este testată într-un studiu clinic care poate fie
să o coroboreze sau să o răstoarne și să conducă la o nouă teorie, care apoi trebuie testată în continuare pentru a vedea dacă prezice mai bine noi observații. Deci există o interacțiune constantă între inferența inductivă (bazată pe observații) și inferența deductivă (bazată pe teorie), până când ne apropiem din ce în ce mai mult de „adevăr”.

Cu toate acestea, există un alt punct în această poveste. Teoriile nu ies doar din fapte. Trebuie să existe un substrat din care să iasă teoria. Poate că acel substrat este o altă teorie anterioară care s-a dovedit a fi inadecvată pentru a explica aceste noi observații și acea teorie, la rândul ei, a înlocuit o teorie anterioară. În orice caz, un aspect al „substratului" este „mintea pregătită" a investigatorului. Dacă investigatorul este cardiolog, de exemplu, el sau ea este instruit să privească fenomenele medicale din perspectiva cardiologiei și cunoaște teoriile precedente și punctele forte și defectele acestora. Dacă cardiologul nu ar fi avut o astfel de pregătire, el sau ea ar fi putut să nu fi văzut legătura. Sau, cu o pregătire diferită, investigatorul ar putea sări la o inferență cu totul diferită. Epidemiologul trebuie să lucreze într-o echipă interdisciplinară pentru a aduce diverse perspective asupra unei probleme și pentru a înrola minți „pregătite" în moduri diferite.

Întrebarea este, cât de bine rezistă o teorie în fața noilor observații? Atunci când multe studii oferă dovezi afirmative în favoarea unei teorii, ne crește asta credința în ea? Dovezile afirmative înseamnă mai multe exemple care sunt în concordanță cu teoria. Dar în ce măsură dovezile susținătoare întăresc o afirmație? Cei care cred că inducția este logica adecvată a științei susțin opinia că dovezile afirmative sunt ceea ce întărește o teorie.

O altă abordare este cea a lui Karl Popper1, poate unul dintre cei mai importanți teoreticieni ai științei. Popper susține că inducția rezultată din acumularea de dovezi afirmative nu întărește o teorie. Inducția, la urma urmei, se bazează pe credința noastră că lucrurile neobservate vor fi ca cele observate sau că viitorul va fi ca trecutul. De exemplu, vedem o mulțime de lebede albe și facem afirmația că toate lebedele sunt albe. Această afirmație este susținută de multe observații. De fiecare dată când vedem o altă lebădă albă, avem mai multe dovezi de susținere. Dar nu putem dovedi că toate lebedele sunt albe, indiferent câte lebede albe vedem.

Pe de altă parte, această afirmație poate fi doborâtă de vederea unei singure lebede negre. Acum ar trebui să ne schimbăm afirmația pentru a spune că majoritatea lebedelor sunt albe și că există unele lebede negre. Această afirmație este probabil mai aproape de adevăr. Cu alte cuvinte, putem respinge afirmația cu un singur exemplu, dar nu o putem dovedi cu multe. (Aserțiunea conform căreia toate lebedele sunt albe este o generalizare descriptivă mai degrabă decât o teorie. O teorie are un înțeles mai bogat care încorporează explicații cauzale și mecanisme care stau la baza. Afirmațiile, precum cele referitoare la culoarea lebedelor, pot fi componente ale unei teorii.)

Potrivit lui Popper, metodologia potrivită este să postulezi o teorie, sau o

conjectură, așa cum o numește el, și să încerci să demonstrezi că este falsă. Cu cât supraviețuiește mai multe astfel de încercări de distrugere, cu atât dovezile sunt mai puternice. Scopul este de a concepe încercări din ce în ce mai agresive de a dărâma afirmația și de a vedea dacă încă supraviețuiește. Dacă nu supraviețuiește unei tentative de falsificare, atunci teoria este renunțată și înlocuită cu alta. El numește aceasta metoda conjecturilor și infirmărilor. Avansul științei către „adevăr" are loc prin eliminarea teoriilor ale căror predicții nu sunt confirmate de observații sau teorii care nu sunt testabile în totalitate , mai degrabă decât prin susținerea teoriilor cu mai multe exemple de unde funcționează. Teoriile științifice utile sunt potențial falsificabile.

Teoriile netestabile sunt acelea în care o varietate de observații contradictorii ar putea fi fiecare în concordanță cu teoria. De exemplu, luați în considerare teoria psihanalitică a lui Freud . Teoria complexului lui Oedip spune că un copil este îndrăgostit de părintele de sex opus. Un băiat își dorește mama și vrea să-și distrugă tatăl. Dacă observăm că un bărbat spune că își iubește mama, asta se potrivește cu teoria. Dacă observăm că un bărbat spune că își urăște mama, asta se potrivește și cu teoria, care ar spune că „formarea reacției" este cea care îl face să-și nege adevăratele sentimente. Cu alte cuvinte, indiferent ce spune omul, nu ar putea falsifica teoria pentru că ar putea fi explicată prin ea. Deoarece nicio observație nu ar putea falsifica teoria lui Oedip, poziția sa ca teorie științifică ar putea fi pusă sub semnul întrebării.

O a treia opinie, și cea mai rezonabilă, este că progresul științei necesită atât inferențe inductive, cât și deductive. Un anumit punct de vedere oferă un cadru pentru observații, care duc la o teorie care prezice noi observații care modifică teoria, ceea ce conduce apoi la observații noi, prezise și așa mai departe, către „adevărul" evaziv, pe care îl vom modifica.
în general, nu ajunge niciodată. A întreba care este primul, teorie sau observație, este ca și cum ai întreba care vine primul, găina sau oul.

În general, progresele în cunoștințe în domeniul sănătății apar prin construirea, testarea și modificarea teoriilor. Epidemiologii fac inferențe inductive pentru a generaliza din multe observații, fac salturi creative ale imaginației pentru a deduce explicații și pentru a construi teorii și folosesc inferențe deductive pentru a testa acele teorii.

Prin urmare, teoriile pot fi folosite pentru a prezice observații. Dar aceste observații nu vor fi întotdeauna exact așa cum le prezicem noi, din cauza erorii și a variabilității inerente a fenomenelor naturale. Dacă observațiile sunt foarte diferite de predicțiile noastre, va trebui să abandonăm sau să modificăm teoria.

Cum testăm gradul de discordanță dintre predicțiile noastre bazate pe teorie și realitatea observațiilor noastre? Testul este un test statistic sau probabilistic. Este testul ipotezei nule, care este piatra de temelie a inferenței statistice și va fi discutată mai târziu. Câteva scrieri clasice excelente despre logica și filosofia științei și aplicațiile în epidemiologie sunt enumerate în secțiunea de referințe de la sfârșitul acestei cărți și, deși unele au fost scrise cu ceva timp în urmă, ele încă pot fi obținute. [2–7]

1.2 Variabilitatea fenomenelor necesită o analiză statistică

Statistica este o metodologie cu domenii largi de aplicare în știință și industrie, precum și în medicină și în multe alte domenii. Un fenomen se poate baza în principal pe un model determinist. Un exemplu este legea lui Boyle , care afirmă că pentru un volum fix o creștere a temperaturii unui gaz determină o creștere a presiunii. De fiecare dată când această lege este testată, apare același rezultat. Singura variabilitate constă în eroarea de măsurare . Multe fenomene din fizică și chimie sunt de o astfel de natură.

Un alt tip de model este un model probabilistic, care presupune că diverse stări ale unui fenomen apar cu anumite probabilități. De exemplu, distribuția inteligenței este în principal probabilistică, adică valorile date ale inteligenței apar cu o anumită probabilitate în populația generală. În biologie, psihologie sau medicină, unde sunt fenomene

influențate de mulți factori care în sine sunt variabili și de alți factori neidentificabili, modelele sunt adesea probabiliste. De fapt, pe măsură ce cunoștințele în fizică au devenit mai rafinate, începe să pară că modelele considerate anterior a fi deterministe sunt probabiliste.

În orice caz, atunci când modelul este în principal probabilistic, sunt necesare tehnici statistice pentru a crește cunoștințele științifice. Prezența variației necesită utilizarea analizei statistice. [7] Când există o variație mică în raport cu un fenomen, se acordă mult mai multă greutate unei cantități mici de dovezi decât atunci când există o mare variație. De exemplu, știm că cancerul pancreatic pare a fi invariabil o boală fatală. Astfel, dacă am găsi un medicament care a vindecat indiscutabil câțiva pacienți de cancer pancreatic, am acorda multă greutate dovezilor că

medicamentul ar reprezenta un leac, mult mai mult decât dacă evoluția acestei boli ar fi mai variabilă. Spre deosebire de acest exemplu, dacă am încerca să stabilim dacă vitamina C vindecă răceala, ar trebui să demonstrăm efectul acesteia la mulți pacienți și ar trebui să folosim metode statistice pentru a face acest lucru, deoarece ființele umane sunt destul de variabile în ceea ce privește raceli. De fapt, în majoritatea fenomenelor biologice și cu atât mai mult în fenomenele sociale și psihologice , există o mare variabilitate.

1.3 Inferență inductivă: Statistica ca tehnologie a metodei științifice

Metodele statistice sunt metode obiective prin care tendințele de grup sunt extrase din observațiile pe mai mulți indivizi separati. Un concept simplu de statistică este calcularea mediilor, procentelor și așa mai departe și prezentarea datelor în tabele și diagrame. Astfel de tehnici de rezumare a datelor sunt într-adevăr foarte importante și esențiale pentru descrierea populației studiate. Cu toate acestea, ele constituie o mică parte din domeniul statisticii. O mare parte a statisticilor implică extragerea de inferențe din eșantioane către o populație cu privire la o caracteristică de interes. Să presupunem că suntem interesați de tensiunea arterială medie a femeilor studente. Dacă am putea măsura tensiunea arterială a fiecărui membru al acestei populații, nu ar trebui să deducem nimic.

Pur și simplu am face o medie a tuturor numerelor pe care le-am obținut. În practică, totuși, luăm un eșantion de studenți (corespunzător selectat), iar pe baza datelor pe care le obținem din eșantion, deducem care ar fi probabil media întregii populații.

Fiabilitatea unor astfel de inferențe sau concluzii poate fi evaluată în termeni de declarații de probabilitate. În raționamentul statistic, deci, facem inferențe inductive, de la particular (eșantion) la general (populație). Astfel, se poate spune că statistica este tehnologia metodei științifice.

1.4 Proiectarea Studiilor

În timp ce generarea de ipoteze poate proveni din observații anecdotice , testarea acelor ipoteze trebuie făcută prin observații controlate, fără părtiniri sistematice. Tehnicile statistice, pentru a fi valide, trebuie aplicate datelor obținute din studii bine concepute. În caz contrar, cunoștințele solide nu sunt avansate.

Există două tipuri de studii: (1) Studii observaționale, în care „Natura" determină cine este expus la factorul de interes și cine nu este expus. Aceste studii demonstrează asocierea. Asocierea poate implica cauzalitate sau poate nu. (2) Studii experimentale, în care anchetatorul determină cine este expus. Acestea pot dovedi cauzalitate.

Studiile observaționale pot avea trei modele de studiu diferite: transversale , caz-control sau prospective . Într-un studiu transversal , măsurătorile sunt luate la un moment dat. De exemplu, într-un studiu transversal al hipertensiunii arteriale și al bolilor coronariene, anchetatorii determină tensiunea arterială și prezența bolilor de inimă în același timp. Dacă găsesc o asociație, nu ar putea spune care a fost primul. Boala de inimă duce la hipertensiune arterială sau hipertensiunea arterială cauzează boli de inimă sau sunt atât hipertensiunea arterială, cât și bolile de inimă rezultatul unei alte cauze comune?

Într-un studiu caz-control al fumatului și al cancerului pulmonar, de exemplu, investigatorul începe cu cazurile de cancer pulmonar și cu controale, iar prin examinarea înregistrărilor sau prin interviuri determină prezența sau absența factorului în care se află. interesat (fumat). Un studiu caz-control este uneori denumit un studiu retrospectiv

deoarece datele privind factorul de interes sunt colectate retroactiv și astfel pot fi supuse diferitelor inexactități.

Într-un studiu prospectiv (sau de cohortă), investigatorul începe cu o cohortă de persoane neboli cu acel factor (adică, cei care fumează) și persoane fără acel factor (nefumători) și merge mai departe într-un timp viitor pentru a determina frecvența dezvoltării boala în cele două grupe. Un studiu prospectiv este cunoscut și sub denumirea de studiu longitudinal. Distincția dintre studiile caz-control și studiile prospective constă în eșantionare. În studiul caz-control, eșantionăm dintre cei bolnavi și nebolnavi, în timp ce într-un studiu prospectiv, eșantionăm dintre cei cu factorul și cei fără factorul . Studiile prospective oferă dovezi mai puternice de cauzalitate decât studiile retrospective, dar sunt adesea mai dificile, mai costisitoare și uneori

imposibil de efectuat, de exemplu, dacă boala studiată durează decenii pentru a se dezvolta sau dacă este foarte rară.

În domeniul sănătății, un studiu experimental pentru a testa o intervenție de un fel se numește studiu clinic. Într-un studiu clinic, investigatorul repartizează pacienții sau participanții la un grup sau altul, de obicei aleatoriu, încercând să mențină toți ceilalți factori constanti sau controlați și compară rezultatul de interes în cele două (sau mai multe) grupuri. Mai multe despre studiile clinice găsiți în Capitolul 6 .

În rezumat, deci, următoarea listă este în ordinea crescătoare a puterii în ceea ce privește demonstrarea cauzalității:

- *Studii transversale:* utile pentru a arăta asocieri, pentru a oferi indicii timpurii asupra etiologiei.

- *Studii caz-control:* utile pentru boli sau afecțiuni rare sau când boala durează foarte mult pentru a se manifesta (alt nume: *studii retrospective*) .

- *Studii de cohortă:* utile pentru furnizarea de dovezi mai puternice ale cauzalității și mai puțin supuse prejudecăților din cauza erorilor de reamintire sau de măsurare (alte denumiri: *studii prospective, studii longitudinale*) .

- *clinice : studii* prospective, experimentale , care oferă cele mai riguroase dovezi de cauzalitate.

1.5 Cum se cuantifică variabilele

Cum testăm o ipoteză? În primul rând, trebuie să stabilim ipoteza într-o manieră cantitativă . Măsura noastră de criteriu trebuie să fie un anumit număr. De exemplu, câți pacienți au murit dintr-un grup de medicamente în comparație cu câți dintre pacienții care nu au primit medicamentul au murit sau care este tensiunea arterială medie a pacienților tratați cu un anumit medicament antihipertensiv în comparație cu tensiunea arterială medie a pacienților care nu asupra acestui medicament. Uneori variabilele sunt greu de cuantificat. De exemplu, dacă evaluați calitatea îngrijirii într-o clinică dintr-un spital în comparație cu clinica altui spital, poate fi uneori dificil să găsiți o măsură cantitativă care să fie reprezentativă pentru calitatea îngrijirii, dar cu toate acestea se poate face și trebuie făcut dacă se dorește

testarea ipotezei.

Există două tipuri de date cu care ne putem ocupa: variabile discrete sau categorice și variabile continue. Variabilele continue, teoretic, pot lua un număr infinit de valori între oricare două puncte fixe. De exemplu, greutatea este o variabilă continuă, la fel ca tensiunea arterială, timpul, inteligența și, în general, variabilele în care pot fi luate măsurători. Variabilele discrete (sau variabilele categorice) sunt variabile care pot lua doar anumite valori numerice fixe. De exemplu, sexul este o variabilă discretă . Îl puteți codifica ca 1 = bărbat, 2 = femeie, dar un individ nu poate avea un cod de 1,5 pentru sex (cel puțin nu teoretic). Variabilele discrete se referă în general la numărare, cum ar fi numărul de pacienți dintr-un anumit grup care trăiesc, numărul de persoane cu o anumită boală și așa mai departe. În capitolul 3 vom lua în considerare o tehnică de testare a unei ipoteze în care variabila este una discretă și, ulterior, vom discuta câteva aspecte ale variabilelor continue , dar mai întâi vom discuta conceptele generale ale testării ipotezelor.

1.6 Ipoteza nulă

Ipoteza pe care o testăm statistic se numește ipoteza nulă. Să luăm un exemplu conceptual simplu. Să presupunem că testăm eficacitatea unui nou medicament la pacienții cu infarct miocardic (atac de cord). Împărțim pacienții în două grupuri - medicament și fără medicament - conform unui design bun

procedurile și utilizarea ca criteriu de măsurare a mortalității în cele două grupuri. Sperăm că medicamentul scade mortalitatea, dar pentru a testa ipoteza statistic , trebuie să o punem într-un fel invers. Spunem că ipoteza noastră este că medicamentul nu face nicio diferență și ceea ce sperăm să facem este să respingem ipoteza „nicio diferență”, pe baza dovezilor din eșantionul nostru de pacienți. Aceasta este cunoscută sub numele de ipoteza nulă . Specificăm ipoteza noastră de testare după cum urmează:

H_0 (ipoteza nulă): Rata mortalității în grupul tratat cu medicamentul A = rata mortalității în grupul tratat cu medicamentul B

Aceasta este echivalentă cu:

H_o: (rata mortalității în grupa A) - (rata mortalității în grupa B) = 0

Testăm acest lucru cu o ipoteză alternativă , cunoscută sub numele de H_A, că diferența în ratele de deces între cele două grupuri nu este egală cu 0.

Apoi colectăm date și notăm diferența observată în mortalitate între grupul A și grupul B. Dacă această diferență observată este suficient de mai mare decât zero, respingem ipoteza nulă. Dacă respingem ipoteza nulă a lipsei de diferență, acceptăm ipoteza alternativă , care este că medicamentul face o diferență.

Când testați o ipoteză, acesta este tipul de raționament pe care îl utilizați:

(1) Voi presupune că ipoteza că nu există nicio diferență este adevărată.
(2) Apoi voi colecta datele și voi observa diferența dintre cele două grupuri.
(3) Dacă ipoteza nulă este adevărată, cât de probabil este ca doar întâmplător să obțin rezultate ca acestea?
(4) Dacă nu este probabil ca aceste rezultate să apară întâmplător în ipoteza că ipoteza nulă este adevărată, atunci voi concluziona că este falsă și voi „accepta" ipoteza alternativă.

1.7 De ce testăm ipoteza nulă?

Să presupunem că credem că medicamentul A este mai bun decât medicamentul B în prevenirea decesului în urma unui atac de cord. De ce nu testăm această credință în mod direct și să vedem care medicament este mai bun decât să testăm ipoteza că medicamentul A este

egal cu medicamentul B? Motivul este că există un număr infinit de moduri în care medicamentul A poate fi mai bun decât medicamentul B, așa că ar trebui să testăm un număr infinit de ipoteze. Dacă medicamentul A provoacă cu 10% mai puține decese decât medicamentul B, este mai bine. Deci mai întâi ar trebui să vedem dacă medicamentul A provoacă cu 10% mai puține decese. Dacă nu provoacă cu 10% mai puține decese, dar dacă provoacă cu 9% mai puține decese, este și mai bine. Apoi ar trebui să testăm dacă observațiile noastre sunt în concordanță cu o diferență de 9% a mortalității între cele două medicamente. Apoi ar trebui să testăm dacă există o diferență de 8% și așa mai departe. Notă: fiecare astfel de ipoteză ar fi înființată ca ipoteză nulă în următoarea formă: medicament A - medicament B mortalitate = 10 %, sau echivalent

(medicamentul A — mortalitatea medicamentului B) — (10 %) = 0 ;
(medicamentul A — mortalitatea medicamentului B) — (9 %) = 0 ;
(medicamentul A — mortalitatea medicamentului B) — (8 %) = 0 ;

Pe de altă parte, atunci când testăm ipoteza nulă fără diferențe, trebuie să testăm doar o singură valoare – o diferență de 0 % – și ne întrebăm dacă observațiile noastre sunt în concordanță cu ipoteza că nu există nicio diferență în ceea ce privește mortalitatea între cele două. droguri. Dacă observațiile sunt în concordanță cu o diferență nulă, atunci nu putem afirma că un medicament este mai bun decât celălalt. Dacă este puțin probabil ca acestea să fie în concordanță cu o diferență nulă, atunci putem respinge acea ipoteză și putem concluziona că există o diferență.

O sursă comună de confuzie apare atunci când investigatorul dorește cu adevărat să arate că un tratament este la fel de bun ca altul (spre deosebire de exemplul de mai sus, în care investigatorul din inima ei crede cu adevărat că un medicament este mai bun). De exemplu, în camera de urgență, este posibil să fi fost concepută o procedură mai rapidă și anchetatorul consideră că poate fi la fel de bună ca procedura standard, care durează mult timp. Tentația într-o astfel de situație este „demonstrarea ipotezei nule". Dar este imposibil să „demonstrezi" ipoteza nulă.

Toate testele statistice pe care le pot face este să respingă ipoteza nulă sau să nu o respingă. Nu demonstrăm ipoteza adunând dovezi afirmative sau susținătoare, pentru că oricâte ori am făcut experimentul și am găsit o diferență aproape de zero, nu am putea fi niciodată siguri că data viitoare când vom face un astfel de experiment, nu vom găsi o diferență uriașă care nu era nici pe departe zero. Este ca exemplul lebedelor albe discutat

mai devreme: indiferent câte lebede albe vedem, nu putem dovedi că toate lebedele sunt albe, deoarece următoarea observare ar putea fi o lebădă neagră. Mai degrabă, încercăm să falsificăm sau să respingem afirmația noastră despre nicio diferență și, dacă afirmația de zero diferență rezistă încercării noastre de respingere, ea supraviețuiește ca o ipoteză în care continuăm să avem credință. Nerespectarea acesteia nu înseamnă că am demonstrat că nu există într-adevăr nicio diferență. Înseamnă pur și simplu că dovezile pe care le avem „sunt în concordanță cu" ipoteza nulă. Rezultatele pe care le-am obținut ar fi putut apărea doar întâmplător dacă ipoteza nulă ar fi fost adevărată. (Poate că designul studiului nostru nu a fost adecvat . Poate că nu am avut destui pacienți.)

Deci, ce se poate face dacă cineva dorește cu adevărat să arate că două

tratamente sunt echivalente? Se poate proiecta un studiu care este suficient de mare pentru a detecta o mică diferență dacă există într-adevăr una. Dacă studiul are puterea (adică o probabilitate mare) de a detecta o diferență care este foarte, foarte, foarte mică și nu reușește să o detecteze, atunci se poate spune cu un grad ridicat de încredere că nu se poate găsi o diferență semnificativă. diferența dintre cele două tratamente. Este imposibil să existe un studiu cu putere suficientă pentru a detecta o diferență de 0 % . Pe măsură ce diferența pe care se dorește să o detecteze se apropie de zero, numărul de subiecți necesar pentru o putere dată se apropie de infinit. Relațiile dintre nivelul de semnificație, putere și dimensiunea eșantionului sunt discutate mai pe larg în capitolul 6 .

1.8 Tipuri de erori

Ideea importantă este că nu putem fi niciodată siguri că avem dreptate fie când acceptăm, fie că respingem o ipoteză. De fapt, riscăm să facem unul dintre cele două tipuri de erori: Putem respinge incorect ipoteza nulă sau de testare, adică putem concluziona că medicamentul reduce mortalitatea atunci când de fapt nu are niciun efect. Aceasta este cunoscută ca eroare de tip I. Sau putem să nu respingem incorect ipoteza nulă sau de testare, adică putem concluziona că medicamentul nu are efect când de fapt reduce mortalitatea. Aceasta este cunoscută sub numele de eroare de tip II. Fiecare dintre aceste erori poartă cu sine anumite consecințe. În unele cazuri, o eroare de tip I poate fi mai gravă; în alte cazuri, o eroare de tip II poate fi mai gravă. Aceste puncte sunt ilustrate în Figura 1.2 .

		ADEVĂRATĂ STARE A NATURII	
		DROGUL NU ARE EFECT H_0 Adevărat	DROGUL ARE EFECT; H_0 Fals, H_A Adevărat
DECIZIE PE BAZĂ DE PROBĂ	NU RESPINGE H_0 Fără efect	NU EROARE	EROARE DE TIP II
	REJECT H_0 (Accept H_A) Efect	TIP I EROARE	FĂRĂ EROARE

Figura 1.2 Testarea ipotezelor și tipurile de erori

Ipoteza nulă (H_o): Medicamentul nu are niciun efect — nicio diferență de mortalitate între pacienții care utilizează droguri și pacienții care nu le folosesc.

Ipoteza alternativă ($HA_)$: Drogul are efect — reduce mortalitatea.

Dacă nu respingem H_o, vom concluziona că nu există nicio relație între droguri și mortalitate. Dacă respingem H_O și acceptăm H_A, ajungem la concluzia că există o relație între drog și mortalitate.

Acțiuni care trebuie întreprinse pe baza deciziei

(1) Dacă credem ipoteza nulă (adică, nu reușim să o respingem), nu vom folosi medicamentul.
Consecințele unei decizii greșite: eroare de tip II. Dacă credem incorect H_o, deoarece în realitate medicamentul este benefic, reținându-l vom permite să moară pacienții care altfel ar fi putut trăi.

(2) Dacă respingem ipoteza nulă în favoarea ipotezei alternative, vom folosi medicamentul.
Consecințele unei decizii greșite: eroare de tip I. Dacă am respins incorect ipoteza nulă, vom folosi medicamentul și pacienții nu beneficiază . Presupunând că medicamentul nu este dăunător în sine, nu rănim direct pacienții, dar din moment ce credem că am găsit leacul, s-ar putea să nu mai testăm alte medicamente.

Nu putem niciodată să cunoaștem în mod absolut „Starea adevărată a naturii”, dar o deducem pe baza unor probe de dovezi .

1.9 Nivel de semnificație și tipuri de eroare

Nu putem elimina riscul de a face unul dintre aceste tipuri de erori, dar putem reduce probabilitățile de a face aceste erori. Probabilitatea de a face o eroare de tip I este cunoscută ca nivelul de semnificație al unui test statistic. Când citiți în literatură că un rezultat a fost semnificativ la nivelul .05, înseamnă că în acest experiment, rezultatele sunt de așa natură încât probabilitatea de a face o eroare de tip I este

mai mică sau egală cu .05. Mai ales în experimente și sondaje, oamenii sunt foarte preocupați de faptul că au o probabilitate scăzută de a face o eroare de tip I și adesea ignoră eroarea de tip II. Aceasta poate fi o greșeală, deoarece în unele cazuri o eroare de tip II este una mai gravă decât o eroare de tip I. În proiectarea unui studiu, dacă doriți să reduceți eroarea de tip I, creșteți automat probabilitatea de eroare de tip II. Pentru a reduce probabilitățile erorii de tip I și de tip II într-un studiu, este necesar să creștem numărul de observații.

Este interesant de observat că regulile Administrației pentru Alimente și Medicamente (FDA) sunt stabilite pentru a reduce probabilitatea de a face erori de tip I. Pentru ca un medicament să fie aprobat pentru comercializare, compania de medicamente trebuie să fie capabilă să demonstreze că nu dăunează și că este eficient. Astfel, multe medicamente sunt respinse deoarece eficacitatea lor nu poate fi demonstrată în mod adecvat. Ipoteza nulă testată este că „acest medicament nu are nicio diferență". Pentru a satisface regulile FDA, această ipoteză trebuie respinsă, probabilitatea de a face o eroare de tip I (adică de a o respinge incorect) fiind destul de mică. Cu alte cuvinte, FDA nu dorește pe piață o mulțime de medicamente inutile. Companiile de medicamente, cu toate acestea, acordă importanță și protecției împotriva erorii de tip II (adică să evite să creadă incorect ipoteza fără diferențe), astfel încât să poată comercializa medicamente potențial benefice.

1.10 Consecințele erorilor de tip I și tip II

Severitatea relativă a acestor erori depinde de situație. Amintiți -vă, o eroare de tip I (cunoscută și sub numele de alfa) înseamnă că spuneți că ceva există într-adevăr (un efect) atunci când de fapt nu există și o eroare de tip II (de asemenea

cunoscută sub numele de eroare beta) înseamnă că îți lipsește ceva care este cu adevărat acolo. Dacă cauți un remediu pentru cancer, o eroare de tip II ar fi destul de gravă. Ți-ar lipsi să găsești tratamente utile. Dacă vă gândiți la un medicament scump pentru a trata o răceală, în mod clar ați dori să evitați o eroare de tip I, adică nu ați dori să faceți afirmații false pentru un remediu pentru răceală.

Este greu de reținut distincția dintre erorile de tip I și II. Poate că această mică pildă ne va ajuta. A fost odată un rege care era foarte gelos pe regina lui. Avea doi cavaleri, Alpha, care era foarte frumos, și Beta, care era foarte urâtă. S-a întâmplat ca Regina să fie îndrăgostită de Beta. Regele, însă, a bănuit că regina avea o aventură cu Alpha și l-a decapitat. Astfel, Regele a făcut ambele tipuri de erori: a bănuit o relație (cu Alpha) unde nu exista și nu a reușit să detecteze o relație (cu Beta) unde chiar exista una. Regina a fugit din regat împreună cu Beta și a trăit fericiți pentru totdeauna, în timp ce Regele a suferit chinuri de vinovăție din cauza respingerii greșite și fatale a lui Alpha.

Mai multe despre alfa, beta, putere și dimensiunea eșantionului apar în capitol 6 . Deoarece testarea ipotezelor se bazează pe probabilități, vom prezenta mai întâi câteva concepte de bază de probabilitate în Capitolul 2 .

Chapter 2
PUTIN PROBABILITATE

Teoria probabilității nu este, în fond, altceva decât bunul simț redus la calcul.

Pierre Simon De Le Place
Theori Analytique des Probabilites (1812–1820)

2.1 Ce este probabilitatea?

Probabilitatea de apariție a unui eveniment este indicată printr-un număr cuprins între 0 și 1. Un eveniment a cărui probabilitate de apariție este 0 este sigur că nu va avea loc, în timp ce un eveniment a cărui probabilitate este 1 este sigur că va avea loc.

Definiția clasică a probabilității este următoarea: dacă un eveniment poate avea loc în N moduri care se exclud reciproc, la fel de probabile și dacă n_A dintre aceste rezultate au atributul A, atunci probabilitatea lui A, scrisă ca P(A), este egală cu n_A / N. Aceasta este o definiție a priori a probabilității, adică se determină probabilitatea unui eveniment înainte ca acesta să se întâmple. Să presupunem că cineva ar trebui să arunce un zar și să știe probabilitatea de a obține un număr divizibil cu trei la aruncarea unui zar. Există șase moduri posibile prin care zarul poate ateriza. Dintre acestea, există două moduri în care numărul de pe fața zarului este divizibil cu trei, un 3 și un 6. Astfel, probabilitatea de a obține un număr divizibil cu trei la aruncarea unui zar este 2/6 sau 1/3.

În multe cazuri, însă, nu suntem capabili să enumerăm toate modalitățile posibile în care poate apărea un eveniment și, prin urmare, folosim definiția frecvenței relative a probabilității. Acesta este definit ca numărul de ori în care evenimentul de interes a avut loc împărțit la numărul total de încercări (sau oportunități pentru ca evenimentul să aibă loc). Deoarece se bazează pe date anterioare, se numește definiția a posteriori a probabilității.

De exemplu, dacă selectați la întâmplare o femeie americană albă, probabilitatea ca ea să moară de boli de inimă este de 0,00199. Acest lucru se bazează pe constatarea că la 100.000 de femei albe americane, 199 au murit din cauza bolilor coronariene (estimările sunt pentru 2011, Sistemul Național de Statistică Vitală, Centrul Național de Statistică în Sănătate, Centrele pentru Controlul și Prevenirea Bolilor).

Când luați în considerare probabilitatea unei femei americane albe cu vârste cuprinse între 45 și 54 de ani, cifra scade la 0,00041 (sau 41 de femei din acel grup de vârstă din 100.000), iar când luați în considerare femeile cu vârsta cuprinsă între 75 și 84 de ani, cifra crește la 0,00913 (sau 913 la 100.000). Pentru bărbații albi 75–84 de ani, este 0,01434 (sau 1.434 la 100.000). Cele două puncte importante sunt (1) pentru a determina o probabilitate, trebuie să specificați populația la care vă referiți, de exemplu, toate femelele albe, masculii albi între 65 și 74 de ani, femelele nealbe între 65 și 74 de ani și așa mai departe; și (2) cifrele de probabilitate sunt revizuite în mod constant pe măsură ce devin disponibile noi date.

Acest lucru ne aduce la noțiunea de frecvență așteptată. Dacă probabilitatea unui eveniment este P și există N încercări (sau oportunități ca evenimentul să se producă), atunci ne putem aștepta ca evenimentul *să* aibă loc *de* $N \times P$ ori. Este necesar să ne amintim că probabilitatea „funcționează" pentru numere mari. Când aruncăm o monedă spunem că probabilitatea ca aceasta să cadă în cap este de 0,50, ne referim la faptul că în multe aruncări, jumătate din timp moneda va ateriza capete. Dacă aruncăm moneda de zece ori, este posibil să obținem trei capete (30 %) sau șase capete (60 %), ceea ce reprezintă o abatere considerabilă de la 50 % la care ne așteptăm. Dar dacă aruncăm moneda de 200.000 de ori, este foarte probabil să fim aproape de a obține exact 100.000 de capete sau 50%.

Frecvența așteptată este într-adevăr modul în care probabilitatea „funcționează". Este dificil de conceptualizat aplicarea probabilității unui individ. De exemplu, când cranicii TV proclamă că vor avea loc, să zicem, 400 de accidente mortale în statul X pe 4 iulie, este imposibil să spunem dacă o persoană individuală va avea într-adevăr un astfel de accident, dar putem fi destul de siguri că numărul de astfel de accidente va fi foarte apropiat de cele 400 prezise (pe baza probabilităților derivate din statisticile anterioare din 4 iulie).

2.2 Combinarea probabilităților

Există două legi pentru combinarea probabilităților care sunt importante. În primul rând, dacă există *evenimente care se exclud reciproc* (adică, dacă unul are loc, celălalt nu poate), probabilitatea ca unul sau celălalt să se producă este *suma* probabilităților lor individuale. Simbolic,

$$P(AorB) = \mathrm{P}(\mathrm{A}) + P(B)$$

Un exemplu în acest sens este următorul: probabilitatea de a obține fie un 3, fie un 4 la aruncarea unui zar este 1/6 + 1/6 = 2/6.

Un lucru util de știut este că suma probabilităților individuale ale tuturor evenimentelor posibile care se exclud reciproc trebuie să fie egală cu 1. De exemplu, dacă A este evenimentul de câștig la o loterie, iar nu A (scris ca A), este evenimentul de a nu câștiga. la loterie, apoi (P (A) + (A) = 1 . 0) și P (A) + (A) = 1 — P (A) .

În al doilea rând, dacă există două evenimente independente (adică, apariția unuia nu este legată de apariția celuilalt), probabilitatea comună ca acestea să se producă împreună (în comun) este produsul probabilităților individuale . Simbolic,

$$P\,(\,A\,\text{și}B\,) = P\,(\,A\,) \times P\,(\,B\,)$$

Un exemplu în acest sens este probabilitatea ca la aruncarea unui zar să obțineți un număr care este atât par, cât și divizibil cu 3. Această probabilitate este egală cu 1/2 × 1/3 = 1/6. (Singurul număr par și divizibil cu 3 este numărul 6.)

Legea probabilității comune este utilizată pentru a testa dacă evenimentele sunt independente . Dacă sunt independenți, produsul probabilităților lor individuale ar trebui să fie egal cu probabilitatea comună. Dacă nu, ei nu sunt independenți. Este baza testului de semnificație chi-pătrat, pe care îl vom lua în considerare în secțiunea următoare.

Să aplicăm aceste concepte la un exemplu medical. Rata mortalității pentru cei cu infarct într-o unitate specială de îngrijire coronariană dintr-un anumit spital este de 15 %. Astfel, probabilitatea ca un pacient cu infarct internat în această unitate coronariană să moară este de 0,15 și să supraviețuiască este de 0,85. Dacă doi bărbați sunt internați în unitatea de îngrijire coronariană într-o anumită zi, fie *A* evenimentul în care primul bărbat moare și fie *B* evenimentul în care al doilea bărbat moare.

Probabilitatea ca ambii să moară este

$$P\,(\,A\,\text{și}\,B\,) = P\,(\,A\,) \times (\,B\,) = .\,15 \times .\,15 = .\,0225$$

Presupunem că aceste evenimente sunt independente unele de altele, așa că le putem multiplica probabilitățile. Rețineți, totuși, că probabilitatea ca unul *sau* altul să moară din cauza atacului de cord *nu este* suma probabilităților lor, deoarece aceste două evenimente nu se exclud reciproc. Este posibil ca ambii să moară

(adică pot apărea atât *A* cât și *B*).

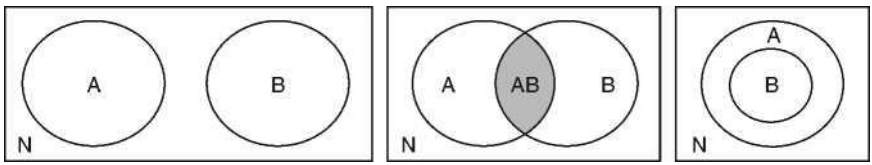


Figura 2.1 Diagramele Venn

Pentru a face acest lucru mai clar, o modalitate bună de abordare a probabilității este prin utilizarea diagramelor Venn, așa cum se arată în Figura 2.1 . Diagramele Venn constau din pătrate care reprezintă universul posibilităților și cercuri care definesc evenimentele de interes.

În diagramele 1, 2 și 3, spațiul din interiorul pătratului reprezintă toate cele N rezultate posibile. Cercul marcat cu A reprezintă toate rezultatele care constituie evenimentul A; cercul marcat cu B reprezintă toate rezultatele care constituie evenimentul B. Diagrama 1 ilustrează două evenimente care se exclud reciproc; un rezultat din cercul A nu poate fi și în cercul B. Diagrama 2 ilustrează două evenimente care pot avea loc împreună: un rezultat din cercul A poate fi și un rezultat aparținând cercului B. Zona umbrită marcată AB reprezintă rezultate care sunt apariția ambelor A și B. Diagrama 3 reprezintă două evenimente în care unul (B) este o submulțime a celuilalt (A); un rezultat din cercul B trebuie să fie, de asemenea, un rezultat care constituie evenimentul A, dar inversul nu este neapărat adevărat.

Se poate observa din diagrama 2 că, dacă vrem ca probabilitatea ca un rezultat să fie fie AorBand dacă adăugăm rezultatele din cercul A la rezultatele din cercul B, am adăugat rezultatele din zona umbrită de două ori. Prin urmare, trebuie să scădem rezultatele din zona umbrită (A și B) scrisă și ca (AB) o dată pentru a ajunge la răspunsul corect. Astfel, obținem rezultatul

$$P\,(A \text{ sau } B) = \mathrm{P}\,(\,\mathrm{A}\,) + P\,(B) - P\,(AB)$$

2.3 Probabilitate condiționată

Acum să luăm în considerare cazul în care șansa ca un anumit eveniment să se întâmple depinde de rezultatul unui alt eveniment. Probabilitatea lui *A* , dat fiind că *B* a avut loc, se numește probabilitate condiționată a

Un dat B și este scris simbolic ca P(A|B). O ilustrare a acestui lucru este oferită de diagrama Venn 2. Când vorbim de probabilitate condiționată, numitorul devine toate rezultatele din cercul B (în loc de toate N rezultate posibile), iar numărătorul este format din acele rezultate care sunt în acea parte a lui A care conține, de asemenea, rezultate aparținând lui B. Aceasta este zona umbrită din diagrama etichetată AB. Dacă ne întoarcem la definiția noastră originală a probabilității, vedem asta

$$P(A|B) \quad \frac{n_{AB}}{n_B}$$

(numărul de rezultate atât în A cât și în B, împărțit la numărul total de rezultate în B).

Dacă împărțim atât numărătorul cât și numitorul la N, numărul total al tuturor rezultatelor posibile, obținem

$$P(A|B) \quad \frac{n_{AB}}{n_B} = \frac{N\ P(A \text{ și } B)}{P(B)}$$

Înmulțirea ambelor părți cu $P(B)$ dă legea multiplicativă *completă* :

$$P(A \text{ și } B) = P(A|B) \times P(B)$$

Desigur, dacă *A* și *B* sunt independenți, atunci probabilitatea lui *A* dat *B* este doar egală cu probabilitatea lui *A* (deoarece apariția lui *B* nu influențează apariția lui *A*) și apoi vedem că

$$P(A \text{ și } B) = P(A) \times P(B)$$

2.4 Probabilitatea Bayesiană

Imaginați-vă că *M* este evenimentul „pierderea memoriei” și *B* este evenimentul „tumoare pe creier”. Putem stabili din cercetările efectuate pe pacienții cu tumori cerebrale probabilitatea *pierderii memoriei în cazul unei tumori pe creier, P(M|B)* . Un clinician, însă, este mai interesat de probabilitatea *unei tumori pe creier, având în vedere că un pacient are pierderi de memorie, P(B|M')* .

Este dificil să se obțină acea probabilitate în mod direct, deoarece ar trebui să se studieze numărul mare de persoane cu pierderi de memorie (care în majoritatea cazurilor provine din alte cauze) și să se determine ce proporție dintre ele au tumori cerebrale.

lui Bayes (sau teorema lui Bayes) estimează P(B | M) . după cum urmează:

$$P(\text{tumoare cerebrală, dată la pierderea memoriei}) = \frac{P(\text{pierderea memoriei, dată fiind o tumoare cerebrală}) \times P(\text{tumoare cerebrală})}{P(\text{pierdere de memorie})}$$

La numitor, evenimentul „pierderii memoriei" poate apărea fie în rândul persoanelor cu tumoră cerebrală, cu probabilitate = P(M | B)P (B) , fie în rândul persoanelor fără tumoră cerebrală, cu probabilitate = P (M | B).)P(B) .

Astfel,

$$P(B \mid M) = \frac{P(M \cap B)P(B)}{P(M \mid B)P(B) + P(M \mid B)P(B)}$$

Probabilitatea generală a unei tumori cerebrale, *P* (*B*), este „probabilitatea a priori ", care este un fel de „cea mai bună presupunere" a prevalenței tumorilor cerebrale.

2.5 Cote și probabilitate

Când se spune că șansele ca un anumit cal *să piardă* o cursă sunt de 4-1, el are o probabilitate de 4/5 = 0,80 de a pierde. Pentru a converti o declarație de cote în capacitatea de probă, adăugăm 4 + 1 pentru a obține numitorul nostru de 5. Cotele ca calul *să câștige* sunt 1–4, ceea ce înseamnă că are o probabilitate de câștig de 1/5 = 0,20:

$$\text{Cotele în favoarea cf } A = \frac{P(A)}{P(\text{not}A)} = \frac{P(A)}{1 - P(A)}$$

$$P(A) = \frac{\text{cote}}{1 + \text{cote}}$$

Cota de a extrage un as = 4 (ași într-un pachet) la 48 (cărți care nu sunt ași) = 1-12; prin urmare, P(as) = 1/13. Cota *împotriva* tragerii unui as = 12-1; P(nu as) = 12/13.

În medicină, cotele sunt adesea folosite pentru a calcula un *raport de cote.* Un raport de cote este pur și simplu raportul dintre două cote. De exemplu, să spunem că într-un anumit studiu care a comparat pacienții cu cancer pulmonar cu martori, s-a constatat că șansele de a fi un caz de cancer pulmonar pentru persoanele care fumează erau 5-4 (5/4) și șansele de a avea cancer pulmonar pentru nefumători. a fost 1–8 (1/8), atunci cota ar fi

$$5 = 4 \_ 5 \times 8\_40$$
$$1 = 8 = 4 \times 1 = T = 0$$

Un raport de șanse de 10 înseamnă că șansele de a fi un caz de cancer pulmonar este de 10 ori mai mare pentru fumători decât pentru nefumători.

Rețineți, totuși, că nu putem determina dintr-o astfel de analiză care este probabilitatea de a face cancer pulmonar pentru fumători, deoarece pentru a face acest lucru ar trebui să știm câți oameni dintre toți fumătorii au dezvoltat cancer pulmonar și nu avem a studiat toți fumătorii; tot ce știm este câți dintre toate cazurile noastre de cancer pulmonar au fost fumători. Nici nu putem obține probabilitatea de cancer pulmonar în rândul nefumătorilor, pentru că ar trebui să ne uităm la o populație de nefumători și să vedem câți dintre ei au dezvoltat cancer pulmonar. Tot ce știm este că fumătorii au șanse de 10 ori mai mari de a avea cancer pulmonar decât nefumătorii.

Mai multe despre acest subiect sunt prezentate în Secțiunea 4.12 .

2.6 Raportul de probabilitate

Un concept înrudit este raportul de probabilitate (LR), care ne spune cât de probabil este ca un anumit rezultat să apară dintr-un set de circumstanțe în raport cu cât de probabil ar apărea rezultatul dintr-un set opus de circumstanțe.

De exemplu, dacă un pacient are o pierdere bruscă a memoriei, am putea dori să cunoaștem probabilitatea acelui simptom pentru o tumoare pe creier, de exemplu. Ceea ce vrem să știm este probabilitatea ca pierderea memoriei să apară din

tumora cerebrală în raport cu probabilitatea ca aceasta să apară dintr-o altă afecțiune. Raportul de probabilitate este un raport al probabilităților condiționate.

$$LR \quad P(\text{pierderea memoriei, dată fiind tumora cerebrală}) \;\_\; P(\text{pierderea memoriei, fără tumori cerebrale}) \;\_\; P(\text{MgivenB}) \;\_\; P(\text{M dat nu B})$$

Desigur, pentru a calcula acest LR, ar trebui să avem estimări ale probabilităților implicate în ecuație, adică ar trebui să știm următoarele: printre persoanele care au tumori cerebrale, ce proporție au pierderi de memorie și printre persoanele care nu Nu au tumori cerebrale, în ce proporție au pierderi de memorie. Poate fi uneori destul de dificil să stabilim numitorul raportului de probabilitate, deoarece ar trebui să cunoaștem prevalența pierderii memoriei în populația generală.

LR este poate mai practic de utilizat decât teorema Bayes, care oferă probabilitatea unei anumite boli având în vedere un anumit simptom. În orice caz, este utilizat pe scară largă într-o varietate de situații, deoarece abordează această întrebare importantă: dacă un pacient prezintă un simptom, care este probabilitatea ca simptomul să se datoreze unei anumite boli, mai degrabă decât unui alt motiv decât această boală?

2.7 Rezumatul probabilității

Dreptul aditiv

$$P(\text{AorB}) \;\_\; P(A) + P(B) - P(\text{AandB})$$

Dacă evenimentele se exclud reciproc: $P(\text{A sau B}) \;\_\; P(A) + P(B)$.

Legea multiplicativă

$$P(\text{A și B}) \;\_\; P(A \mid B) \times P(B)$$

Dacă evenimentele sunt independente: $P(\text{A și B}) \;\_\; P(A) \times P(B)$.

Probabilitate condiționată

$$P(A|B) \qquad \frac{P(AandB)}{P(B}$$

Bayes' Theorem

$$P(B|A) \qquad \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B)P(B)}$$

Odds of A

$$\frac{P(A)}{1 - P(A)}$$

Likelihood Ratio

$$\frac{P(A \quad Bl}{P^{\wedge}A \quad B)}$$

Chapter 3
MAI MULT DESPRE STATISTICĂ

> Un statistician este cineva care, cu capul într-un cuptor și picioarele într-o găleată cu apă cu gheață, la întrebarea cum se simte, răspunde: „În medie , mă simt bine".
>
> Anonim

Diferite tehnici statistice sunt adecvate în funcție de faptul că variabilele de interes sunt discrete sau continue. Vom lua în considerare mai întâi cazul variabilelor discrete și vom prezenta testul chi-pătrat, iar apoi vom discuta metode aplicabile variabilelor continue.

3.1 Chi-Pătrat pentru 2 × 2 Mese

Testul chi-pătrat este o metodă statistică pentru a determina dacă rezultatele unui experiment pot apărea întâmplător sau nu. Să luăm, așadar, exemplul testării unui medicament anticoagulant pe paciente de sex feminin cu infarct miocardic. Sperăm că medicamentul scade mortalitatea, dar am stabilit ipoteza noastră nulă după cum urmează:

- Nulă ipoteză — Nu există nicio diferență în ceea ce privește mortalitatea între grupul de pacienţi tratat şi grupul de control
- Alternativ ipoteză — Mortalitatea în grupul tratat este mai mică decât în grupul de control

(Datele pentru exemplul nostru provin dintr-un studiu realizat cu mult timp în urmă și se referă la un grup specific cu risc ridicat. [8] Sunt utilizate în scopuri ilustrative și nu reflectă ratele actuale de mortalitate pentru persoanele cu infarct miocardic.)

Apoi ne înregistrăm datele într-un tabel *de contingență 2 × 2* în care fiecare

pacient este clasificat ca aparținând uneia dintre cele patru celule:

Frecvențele observate

	Controla	Tratat	
A trăit	89	223	312
A murit	40	39	79
Total	129	262	391

Mortalitatea la lotul martor este de 40/129 = 31 %, iar la cel tratat este de 39/262 = 15 %. Dar ar fi putut această diferență să fi apărut întâmplător ? Folosim testul chi-pătrat pentru a răspunde la această întrebare. Ceea ce ne întrebăm cu adevărat este dacă cele două categorii de clasificare (control vs. tratat prin trăit vs. murit) sunt independente una de cealaltă. Dacă sunt independente, la ce frecvențe ne-am aștepta în fiecare dintre celule? Și cât de diferite sunt frecvențele noastre observate de cele așteptate? Cum măsurăm dimensiunea diferenței?

Pentru a determina frecvențele așteptate, luați în considerare următoarele:

	Control	Treated	
Lived	*a*	*b*	*(a + b)*
Died	*c*	*d*	*(c + d)*
Total	*(a + c)*	*(b + d)*	*N*

Dacă categoriile sunt independente, atunci probabilitatea ca un pacient să fie atât control , cât și în viață este P(control) × P(trăit). [Aici aplicăm legea la care se face referire în capitolul 2 pe probabilitatea comună a două evenimente independente.]

Frecvența așteptată a unui eveniment este egală cu probabilitatea evenimentului înmulțit cu numărul de studii = *N* × *P*. Deci numărul așteptat de pacienți care sunt *atât* martori, *cât și* vii este

$$N \times P\{\text{controlând trăit}) = N \times P(\text{control}) \times P(\text{Iived})$$

$$N(a+c) \times (a+b)\text{I } NN = (a + \text{c}) \times \frac{(a + b)}{N}$$

În cazul nostru, rezultă următorul tabel:

	Controla	Tratat

A trăit	129	$\times \frac{312}{391} =$ 103	262	$\times \frac{312}{391} =$ 209	312
A murit	129	$\times \frac{79}{391} =$ 26	262	$\times \frac{79}{391} =$ 53	79
Total		129		262	391

Un alt mod de a privi acest lucru este de a spune că, deoarece 80 % dintre pacienții din totalul studiului au trăit (adică, 312/391 = 80 %), ne-am aștepta ca 80 % dintre pacienții de control și 80 % dintre pacienții tratați să trăi. Aceste așteptări diferă, după cum vedem, de frecvențele observate notate mai devreme, adică acei pacienți tratați au avut, de fapt, o mortalitate mai mică decât cei din grupul de control.

Ei bine, acum că avem un tabel cu frecvențe observate și un tabel cu valori așteptate, de unde știm cât de diferite sunt acestea? Diferă ele doar întâmplător sau există vreun alt factor care le face să difere? Pentru a determina acest lucru, calculăm o valoare numită chi-pătrat (scrisă și ca χ^2). Aceasta se obține prin luarea valorii observate în fiecare celulă, scăzând din aceasta valoarea așteptată în fiecare celulă, pătrarea acestei diferențe și împărțirea la valoarea așteptată pentru fiecare celulă. Când se face acest lucru pentru fiecare celulă, cele patru cantități rezultate sunt adăugate împreună pentru a da un număr numit chi-pătrat. Din punct de vedere simbolic, această formulă este următoarea:

$$\frac{(O_a - e_a)^2}{e_a} + \frac{(O_b - e_b)^2}{e_b} + \frac{(O_c - e_c)^2}{e_c} + \frac{(O_d - e_d)^2}{e_d}$$

unde O este frecvența observată și e este frecvența așteptată în fiecare celulă.

Acest număr, numit chi-pătrat, este o statistică care are o distribuție cunoscută . Ceea ce înseamnă, în esență, este că pentru un număr infinit de astfel de tabele 2 X 2, s-au calculat chi-pătrat și astfel știm care este probabilitatea de a obține anumite valori ale chi-pătratului. Astfel, când calculăm un chi-pătrat pentru un anumit tabel de contingență 2 x 2, știm cât de probabil este să fi putut obține o valoare la fel de mare ca cea care

am obținut de fapt strict întâmplător, sub ipoteza independenței este cea corectă, adică dacă cele două categorii de clasificare nu erau legate între ele sau dacă ipoteza nulă ar fi adevărată. Valoarea specială a chi-pătratului pe care o obținem

pentru exemplul nostru se întâmplă să fie 13,94.

Din cunoștințele noastre despre distribuția valorilor chi-pătratului, știm că dacă ipoteza noastră nulă este adevărată, adică dacă nu există nicio diferență de mortalitate între grupul de control și cel tratat, atunci probabilitatea ca să obținem o valoare a chi-ului. -pătrat la fel de mare sau mai mare decât 13,94 doar întâmplător este foarte, foarte scăzut; de fapt această probabilitate este mai mică de .005. Deoarece nu este probabil să obținem o valoare atât de mare a chi-pătratului din întâmplare în ipoteza noastră nulă, trebuie să fie că aceasta a apărut nu întâmplător, ci pentru că ipoteza noastră nulă este incorectă. Prin urmare, respingem ipoteza nulă la nivelul de semnificație .005 și acceptăm ipoteza alternativă, adică concluzionăm că în rândul femeilor cu infarct miocardic, noul medicament reduce mortalitatea. Probabilitatea de a obține aceste rezultate numai întâmplător este mai mică de 5/1000 (.005). Prin urmare , probabilitatea de respingere a ipotezei nule, atunci când aceasta este de fapt adevărată (eroare de tip I) este mai mică de .005.

Probabilitățile de obținere a diferitelor valori ale chi-pătratului sunt tabelate în majoritatea textelor de statistică standard, astfel încât procedura este de a calcula valoarea chi-pătratului și apoi de a o căuta în tabel pentru a determina dacă este sau nu semnificativă. Acea valoare a chi-pătratului care trebuie obținută din date pentru a fi semnificativă se numește valoare critică. Valoarea critică a chi-pătratului la nivelul de semnificație de 0,05 pentru un tabel 2 × 2 este 3,84. Aceasta înseamnă că atunci când obținem o valoare de 3,84 *sau mai mare* dintr-un tabel 2 × 2, putem spune că există o diferență semnificativă între cele două grupuri. Anexa 1 oferă câteva valori critice pentru chi-pătrat și pentru alte teste.

În utilizarea efectivă , se aplică o corecție pentru tabele 2 × 2 cunoscute sub numele de corecție Yates , iar calculul se face folosind formula

$$N[\,|\,ad - bc\,|\,-N]^2$$

$$(a+b)(c+d)(a+c)(b+d)$$

Notă: |ad-bc| înseamnă valoarea absolută a diferenței dintre *a* × *d* și b × *c*. Cu alte cuvinte, dacă *a* × *d* este mai mare decât *b* × *c*, scădem *bc* din *ad* ; dacă *bc* este mai mare decât *ad* , scădeți *ad* din *bc*. Chi-pătratul corectat astfel calculat este 12,95, încă cu mult peste 3,84 necesar pentru semnificație.

Testul chi-pătrat nu trebuie utilizat dacă numerele din celule sunt prea mici. Regulile generale: Când *N total* este mai mare de 40, utilizați testul chi-pătrat cu corecția lui Yates . Când *N* este între 20 și 40 și frecvența așteptată în fiecare

dintre cele patru celule este de 5 sau mai mult, utilizați testul chi-pătrat corectat. Dacă cea mai mică frecvență așteptată este mai mică de 5 sau dacă *N* este mai mică de 20, utilizați testul Fisher .

În timp ce testul chi-pătrat aproximează probabilitatea, testul exact al lui Fisher oferă probabilitatea exactă de a obține un tabel cu valori ca cele obținute sau chiar mai extreme. Un exemplu de calcul este prezentat în Anexa 2 . Calculele sunt grele, dar testul exact al lui Fisher este de obicei inclus în majoritatea programelor de statistică pentru computerele personale. Mai multe despre acest subiect pot fi găsite în cartea *Statistical Methods for Rates and Proportions* de Joseph L. Fleiss. Important este să știi când testul chi-pătrat este sau nu potrivit.

3.2 Testul McNemar

Să presupunem că avem situația în care se fac măsurători pe același grup de oameni înainte și după o intervenție sau să presupunem că suntem interesați de acordul dintre doi judecători care evaluează același grup de pacienți asupra unor caracteristici. În astfel de situații, măsurile înainte și după, sau opiniile a doi judecători, nu sunt independente unul de celălalt, deoarece aparțin acelorași indivizi. Prin urmare, testul chi-pătrat sau testul exact al lui Fisher nu este adecvat. În schimb, putem folosi testul McNemar.

Luați în considerare următorul exemplu. Cazurile pacienților care au fost suspectați de boală cardiacă ischemică (o scădere a fluxului sanguin către inimă din cauza înfundarii arterelor) au fost prezentate la doi experți în cardiologie. Medicii au fost rugați să își dea o opinie pe baza informațiilor disponibile despre pacient. Ar putea

TABELUL 3.1

		EXPERT 1: Medical	EXPERT 1: Surgical	
EXPERT 2	Medical	$a = 397$	$b = 97$	$a + b = 494$
	Surgical	$c = 91$	$d = 76$	$c + d = 167$
		$a + c = 488$	$b + d = 173$	$N = 661$

recomandă fie (1) ca pacientul să urmeze terapie medicală, fie (2) ca pacientul să aibă o angiografie, care este un test invaziv, pentru a determina dacă pacientul este un candidat potrivit pentru intervenția chirurgicală de bypass de arteră coronariană (cunoscută ca CABG) . Tabelul 3.1 prezintă rezultatele acestor judecăți pe 661 de pacienți.

Rețineți că în celula b Expertul 1 a sfătuit intervenția chirurgicală și Expertul 2 a sfătuit terapie medicală pentru 97 de pacienți, în timp ce în celula c Expertul 1 a sfătuit terapie medicală și Expertul 2 a sfătuit intervenția chirurgicală pentru 91 dintre pacienți. Astfel, cei doi medici nu sunt de acord în 188 din cele 661 de cazuri sau 28 % din timp. Celulele a și d reprezintă pacienți despre care cei doi medici sunt de acord. Ei sunt de acord în 473 din 661 de cazuri sau 72 % din cazuri.

Pentru a determina dacă dezacordul observat ar fi putut apărea doar întâmplător sub ipoteza nulă a lipsei unui dezacord real în recomandările dintre cei doi experți, calculăm un tip de valoare chi-pătrat după cum urmează:

$$\chi^2 (\textit{chi-pătrat}) = \frac{2 \cdot s}{(b + c)}; \quad \frac{(|b - c| - 1|)^2}{188} = \frac{25}{} = :13$$

(| bc | înseamnă valoarea absolută a diferenței dintre cele două celule, adică indiferent de semn; — 1 din numărător este analog cu corecția lui Yates pentru chi-pătrat descrisă în secțiunea 3.1 și oferă o mai bună

aproximarea distribuției chi-pătrat.) Un chi-pătrat de 0,13 nu atinge valoarea critică a chi-pătrat de 3,84 necesară pentru un nivel de semnificație de 0,05, așa cum este descris în secțiunea 3.1 , deci nu putem respinge ipoteza nulă și concluzionăm că datele noastre sunt în concordanță fără nicio diferență în opiniile celor doi experți. Dacă testul chi-pătrat semnificativ, ar trebui să respingem ipoteza nulă și să spunem că experții nu sunt în mod semnificativ de acord. Cu toate acestea, un astfel de test nu ne spune despre puterea acordului lor, care poate fi evaluat printr-o statistică numită Kappa.

3.3 Kappa

Cei doi experți ar putea fi de acord doar întâmplător, deoarece ambii experți sunt mai predispuși să recomande terapie medicală pentru acești pacienți. Kappa este o statistică care ne spune amploarea acordului dintre cei doi experți dincolo de acordul întâmplător.

Proporția acordului observat — Proporția acordului întâmplător
1 — *Proportion of agreement by chance*

Pentru a calcula numărul așteptat de cazuri în fiecare celulă a tabelului, urmează procedura descrisă pentru chi-pătrat în secțiunea 3.1 . Celulele *a* și *d* din tabelul 3.1 reprezintă acordul. Numărul așteptat doar întâmplător este

$$\text{celula } a: \frac{494 \times 488}{661} = 365$$

$$\text{celula } d: \frac{167 \times 173}{661} = 44$$

Deci proporția de acord așteptată doar din întâmplare este

$$\frac{365 + 44}{661} = .619$$

adică, doar întâmplător, experții ar fi de așteptat să fie de acord în 62 % din cazuri. Proporția acordului observat este

$$\frac{397 + 76}{661} \quad :716$$

$$Kappa = \frac{:716 - :619}{1 - :619} \quad \frac{:097}{:381} = .25$$

Dacă cei doi experți ar fi de acord doar la nivel de șansă, Kappa ar fi 0; dacă cei doi experți ar fi de acord perfect, Kappa ar fi 1.

3.4 Descrierea unei populații: Utilizarea abaterii standard

În cazul variabilelor continue, ca și în cazul variabilelor discrete, putem fi interesați de descriere sau de inferență. Când dorim să descriem o populație cu privire la o anumită caracteristică, folosim în general media sau media ca indice de *tendință centrală* a datelor.

Alte măsuri de tendință centrală sunt *mediana* și *modul* . Mediana este acea valoare peste care se află 50 % din celelalte valori și sub care se află 50 % din valori. Este valoarea medie sau a 50-a percentila. Pentru a găsi mediana unui set de scoruri, le aranjam în ordine crescătoare (sau descrescătoare) și găsim valoarea din mijloc dacă există un număr impar de scoruri, sau media dintre cele două scoruri mijlocii dacă există un număr par de scoruri. . Modul este valoarea care apare cu cea mai mare frecvență. Pot exista mai multe moduri într-un set de scoruri, dar o singură valoare mediană și o valoare medie. Aceste definiții sunt ilustrate mai jos. Media este măsura tendinței centrale cel mai des folosită în statisticile inferențiale.

Măsuri de tendință centrală

Set de scoruri	Ordonat
12	6
12	8
6	10
8	11 Mediană
11	12 Modul

10	12
15	15
SUMA: 74	*Medie* ¼ 74/7 = 10,6

Adevărata medie a populației se numește m și o estimăm din datele obținute dintr-un eșantion al populației. Media eșantionului se numește x̄ (se citește ca x bară). Trebuie să avem grijă să precizăm exact populația din care prelevăm un eșantion. De exemplu, în populația generală, IQ-ul mediu este de 100, dar IQ-ul mediu al populației de copii de 6-11 ani ai căror tați sunt absolvenți de facultate este de 112. [9] Prin urmare, dacă luăm un eșantion din oricare dintre aceste populații, am estima o medie a populației diferită și trebuie să specificăm la care populație facem inferențe.

Cu toate acestea, media nu oferă o descriere adecvată a unei populații. De asemenea, este nevoie de o măsură a variabilității datelor în jurul valorii medii. Două grupuri pot avea aceeași medie, dar pot fi foarte diferite. De exemplu, luați în considerare un grup ipotetic de copii al căror IQ individual este 100; astfel, media este 100. Comparați aceasta cu un alt grup a cărui medie este de asemenea 100, dar include indivizi cu IQ de 60 și cei cu IQ de 140. Trebuie făcute diferite afirmații despre aceste două grupuri: unul este compus din toți indivizii medii și altele includ atât retardații, cât și genii.

Cel mai frecvent utilizat indice de variabilitate este deviația standard (sd), care este un tip de măsură legat de distanța medie a scorurilor față de valoarea medie a acestora. Pătratul abaterii standard se numește varianță. Abaterea standard a populației este notată cu litera greacă σ (sigma). Când este calculat dintr-un eșantion, este scris ca sd și este ilustrat în exemplul de mai jos:

Scoruri IQ		Abateri de la medie		Scoruri la pătrat pentru B
Grupa A	Grupa B	$X_i - X_B$	$(X_i - X_B)^2$	$2\ x\ b^2$
100	60	— 40	1.600	3.600
100	140	40	1.600	19.600
100	80	— 20	400	6.400
100	120	20	400	14.400
$\Sigma = 400$ X_A = medie = 100	$\Sigma = 400$ X_B = medie = 100	$\Sigma = 0$	Σ = 4.000 de abateri la pătrat	Σ = 44.000 suma pătratelor
Notă: simbolul „Σ ” înseamnă „suma”. Notă: Suma abaterilor de la medie, ca în coloana 3, este întotdeauna 0; de aceea însumăm abaterile pătrate, ca în coloana 4.				

x_A = *medie* = ∫ = 100 ; x_B = ∫ = 100

$$:d: = \sqrt{\frac{\Sigma a \text{ (fiecare valoare} - \text{media grupului)}}{n - 1}} S = \sqrt{\frac{\Sigma (X - X)^2}{n - 1}}$$

$S:d:_A = 3 = 0;$

(În grupa A, deoarece fiecare scor este egal cu media lui 100, nu există abateri de la media lui A.)

$$s:d:_B = \sqrt{\frac{4.000}{\text{ffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffi}}} = \sqrt{\Gamma 333} = 36:51$$

O formulă echivalentă pentru sd care este mai potrivită pentru calculele reale este

$$s:d: - \sqrt{\frac{\text{-ɔ}}{-1}\ nx^2}$$

Pentru grupa B avem

$$Sd: = \sqrt{\frac{44{,}000 - 4(100)^2}{3}} = \sqrt{\frac{44{,}000}{3}} = \sqrt{\frac{40{,}000}{3}} = \sqrt{\frac{4{,}000}{1}} = 36{:}51$$

$$Variance = (sd)^2$$

Rețineți că media ambelor grupuri este 100, dar abaterea standard a grupului A este 0, în timp ce sd a grupului B este 36,51. (Împărțim abaterile pătrate cu $n-1$ mai degrabă decât cu n , deoarece estimăm populația σ din datele eșantionului, iar împărțirea la $n-1$ oferă o estimare mai bună. Motivul matematic este complex și dincolo de scopul acestei cărți.)

3.5 Semnificația abaterii standard: distribuția normală

Abaterea standard este o măsură a dispersiei sau răspândirii datelor. Luați în considerare o variabilă precum IQ, care este distribuită în mod normal, adică poate fi descrisă prin curba familiară, în formă de clopot, în care majoritatea valorilor se încadrează în jurul mediei cu un număr descrescător de valori la ambele extreme. Într-un astfel de caz, 68 % dintre valori se află în limita unei abateri standard de ambele părți ale mediei, 95 % dintre valori se află la 2 abateri standard ale mediei și 99 % dintre valori se află la 3 abateri standard ale mediei. . (Testul IQ a fost construit inițial astfel încât să aibă o medie de 100 și o abatere standard de 16.)

Acest lucru este ilustrat în Figura 3.1 .

În populația generală, 95 % dintre oameni au un IQ între 68 și 132. Aproximativ 2,5 % dintre oameni au un IQ peste 132 și alți 2,5 % dintre oameni au un IQ sub 68. (Acest lucru este indicat de zonele umbrite de la cozile curbe.)

Dacă estimăm dintr-un eșantion și dacă există un număr mare de observații, abaterea standard poate fi estimată din intervalul *datelor* , adică diferența dintre cea mai mică și cea mai mare valoare. Împărțirea intervalului la 6 oferă o estimare aproximativă a abaterii standard dacă distribuția este normală, deoarece 6 abateri standard (3 de fiecare parte a mediei) cuprind 99 % sau aproape toate datele.

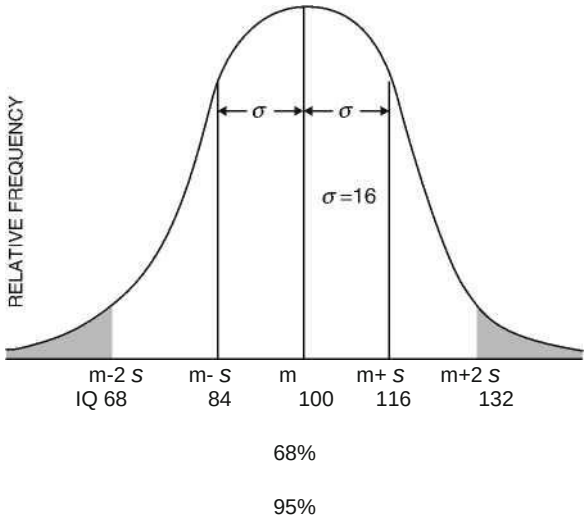


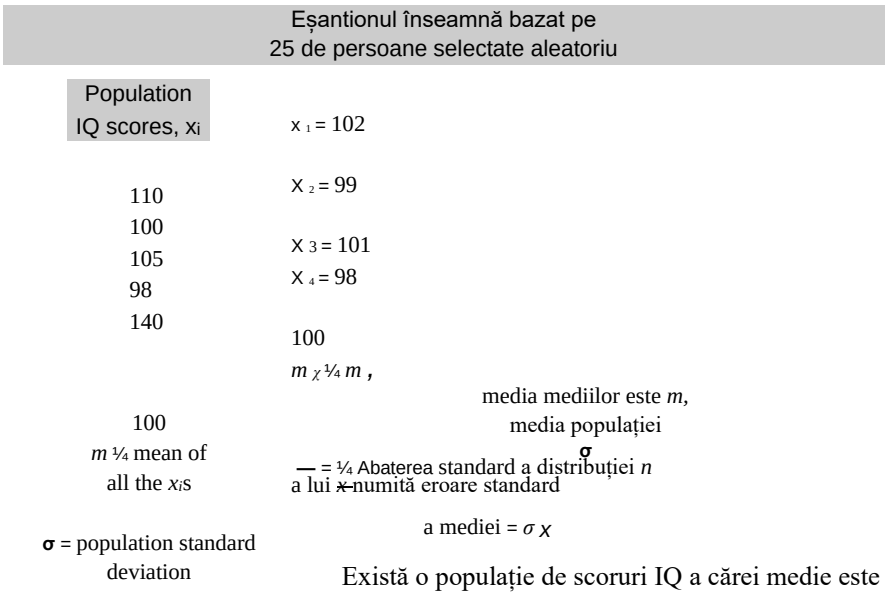
Figura 3.1 Distribuție normală

La nivel individual, clinic, cunoașterea abaterii standard este foarte utilă pentru a decide dacă o constatare de laborator este normală, în sensul „sănătos”. În general, o valoare care este la mai mult de 2 abateri standard de la medie este suspectă și poate că trebuie efectuate teste suplimentare.

De exemplu, să presupunem că, ca medic, vă confruntați cu un bărbat adult care are o valoare a hematocritului de 39. Hematocritul este o măsură a cantității de globule roșii într-o cantitate măsurată de sânge. Un hematocrit scăzut poate implica anemie, care la rândul său poate implica o afecțiune mai gravă. De asemenea, știți că valoarea medie a hematocritului la bărbații adulți este de 47. Știți dacă pacientul cu o citire de 39 este normal (în sensul sănătății) sau anormal? Trebuie să cunoașteți abaterea standard a distribuției hematocritelor la oameni înainte de a putea determina dacă 39 este o valoare normală. De fapt, abaterea standard este de aproximativ 3,5; astfel, plus sau minus 2 abateri standard în jurul mediei rezultă în intervalul de la 40 la 54, astfel încât 39 ar fi ușor scăzut. Pentru femelele adulte, hematocritul mediu este de 42 cu o abatere standard de 2,5, astfel încât intervalul de plus sau minus 2 abateri standard de la medie este de la 37 la 47. Astfel, dacă o femeie adultă a venit la dumneavoastră cu o citire a hematocritului de 39, ea ar fi considerată în intervalul „normal”.

3.6 Diferența dintre abaterea standard și eroarea standard

Adesea, datele din literatură sunt raportate ca ± sd (citite ca medie+ sau — 1 abatere standard). Alte ori sunt raportate ca ± se (se citesc ca medie + sau —1 eroare standard). *Eroarea standard* și *abaterea standard* sunt adesea confundate, dar au funcții destul de diferite. Pentru a înțelege conceptul de eroare standard, trebuie să vă amintiți că scopul statisticilor este de a trage inferențe din eșantioane de date către populația din care provin aceste eșantioane. Mai exact, ne interesează estimarea mediei adevărate a unei populații pentru care avem o medie a eșantionului bazată, de exemplu, pe 25 de cazuri. Imaginați-vă următoarele:



Există o populație de scoruri IQ a cărei medie este 100 și abaterea sa standard este 16. Acum imaginați-vă că extragem la întâmplare un eșantion de 25 de persoane din acea populație și calculăm media eșantionului ⋆ . Această medie a eșantionului se întâmplă să fie 102. Dacă am lua un alt eșantion de 25 de indivizi , probabil că am obține o medie a eșantionului ușor diferită, de exemplu , 99. Să presupunem că am făcut acest lucru în mod repetat de un număr infinit (sau foarte mare) de ori , de fiecare dată când aruncam proba, ne-am tras înapoi în

Distribuția lui X

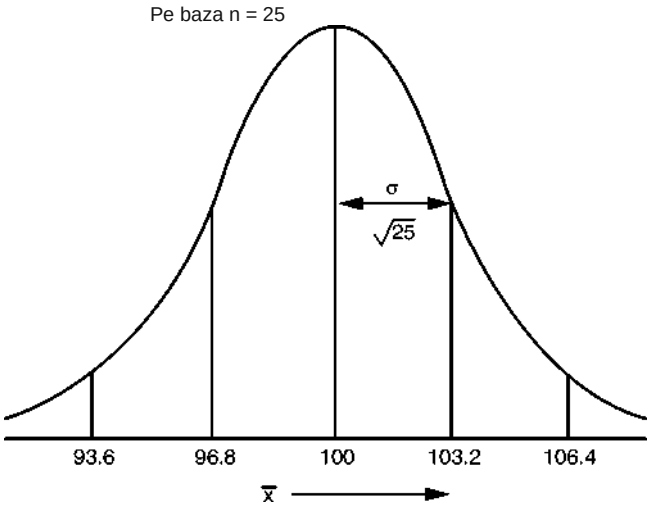


Figura 3.2 Distribuția mediilor eșantionului

populație din care am eșantiona din nou 25 de persoane. Am avea atunci un număr foarte mare de astfel de mijloace de eșantionare. Aceste mijloace eșantionului ar forma o distribuție normală. Unele dintre ele ar fi foarte aproape de media reală a populației de 100, iar unele s-ar afla la fiecare capăt al acestei „distribuții a mijloacelor", ca în Figura 3.2 .

Această distribuție a mediilor eșantionului ar avea propria sa abatere standard, adică o măsură a răspândirii datelor în jurul mediei datelor. În acest caz, datele sunt medii eșantionare mai degrabă decât valori individuale. Abaterea standard a acestei distribuții a mediilor se numește eroarea standard a mediei.

Trebuie subliniat că această distribuție a mijloacelor, care este numită și distribuția prin eșantionare a mijloacelor, este un construct teoretic. Evident, nu ocolim măsurarea eșantioanelor populației pentru a construi o astfel de distribuție. De obicei, de fapt, luăm doar un eșantion de 25 de persoane și ne imaginăm care ar putea fi această distribuție. Cu toate acestea, datorită anumitor derivări matematice, știm multe despre această distribuție teoretică a mijloacelor populației și, prin urmare, putem desena importante

inferențe bazate pe o singură medie eșantion. Ceea ce știm este că distribuția mediilor este o distribuție normală, că media ei este aceeași cu media populației valorilor individuale, adică media mediilor este m și că abaterea sa standard este egală cu abaterea standard a valorilor individuale originale împărțite la rădăcina pătrată a numărului de persoane din eșantion.

standard a mediei =

$$\sigma_{\chi} = \frac{\sigma}{n}$$

In acest caz ar fi

$$\frac{16}{ffi} \quad \frac{16}{5} = 3.2$$

Distribuția mijloacelor ar arăta așa cum se arată în Figura 3.2 .

Vă rugăm să rețineți că atunci când vorbim despre valorile populației, pe care de obicei nu le știm, dar încercăm să le estimăm, ne referim la medie ca *m* și la abaterea standard ca σ . Când vorbim despre valorile calculate din probe, ne referim la medie ca ~~x~~, abaterea standard ca sd și eroarea standard ca se

Acum imaginați-vă că avem o distribuție a mediilor bazată pe eșantioane de 64 de indivizi. Media acestor mijloace este, de asemenea, *m* , dar dispersia sa, sau eroarea standard , este mai mică. Este 16 / ^64 = 16 / 8 = 2. Acest lucru este ilustrat în Figura 3.3 .

Se vede cu ușurință că, dacă luăm un eșantion de 25 de indivizi, este probabil ca media lor să fie mai aproape de media adevărată decât valoarea unui singur individ, iar dacă luăm un eșantion de 64 de indivizi, este probabil ca media lor să fie și mai aproape de adevărata medie. medie decât a fost media pe care am obținut-o din eșantionul de 25. Astfel, cu cât dimensiunea eșantionului este mai mare, cu atât estimarea noastră a mediei adevărate a populației este mai bună.

Abaterea standard este utilizată pentru a descrie dispersia sau variabilitatea scorurilor. Eroarea standard este utilizată pentru a deduce inferențe despre media populației din care avem un eșantion. Tragem astfel de inferențe construind intervale de încredere, care sunt discutate în secțiune 3.11 .

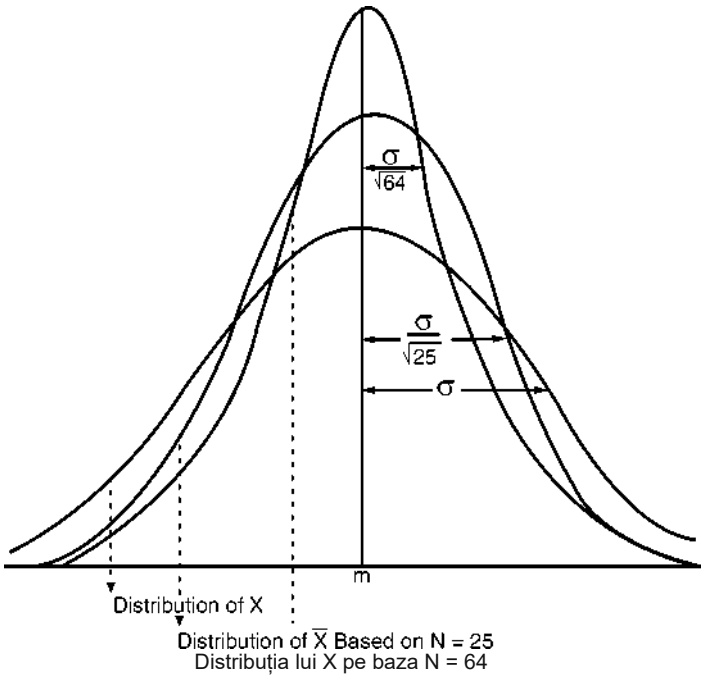


Figura 3.3 Distribuția mediilor pentru diferite dimensiuni ale eșantionului

3.7 Eroare standard a diferenței dintre două mijloace

Acest concept este analog cu conceptul de eroare standard a mediei. Eroarea standard a diferențelor dintre două medii este abaterea standard a unei distribuții teoretice a diferențelor între două medii. Imaginați-vă un grup de bărbați și un grup de femei, fiecare având o măsurătoare IQ. Să presupunem că luăm un eșantion de 64 de bărbați și un eșantion de 64 de femei, calculăm IQ-urile medii ale acestor două eșantioane și obținem diferențele lor. Dacă ar fi să facem acest lucru de un număr infinit de ori, am obține o distribuție a diferențelor între mediile eșantionului a două grupuri de 64 fiecare. Aceste scoruri de diferență ar fi distribuite în mod normal; media lor ar fi adevărata diferență medie dintre populațiile de bărbați și femei (pe care încercăm să o deducem din eșantioane) și

abaterea standard a acestei distribuții se numește eroarea standard a diferențelor dintre două medii.

Eroarea standard a diferenței dintre două medii ale populațiilor X și Y este dată de formula

$$\sigma_{x-y} \qquad \sigma^2 \sigma_y^2 \qquad x$$
$$n_x \qquad n_y$$

unde σ_x^2 este varianța populației X și σ_y^2 este varianța populației Y, n_x este numărul de cazuri din eșantionul din populația X și n_y este numărul de cazuri din eșantionul din populația Y.

În unele cazuri, știm sau presupunem că variațiile celor două populații sunt egale între ele și că variațiile pe care le calculăm din eșantioanele pe care le-am extras sunt ambele estimări ale unei variații comune . Într-o astfel de situație, am dori să punem în comun aceste estimări pentru a obține o estimare mai bună a varianței comune. Notăm această estimare comună ca $s_{pooled} = s_p$ și we e calculați eroarea standard a diferenței dintre medii ca

$$s.e._{x-y} = \sqrt{s_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y}\right)} = s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Calculăm s^2_p din eșantion de date:

$$s2_p \qquad (n_x - l)s^2 + (n_y - 1)s^2$$
$$n_\chi + \Pi_y - 2$$

Acesta este echivalentul cu

$$s^2_p \qquad \Sigma(x_i - x)^2 + (y_i - y)^2$$
$$\prod x + \prod y - 2$$

Deoarece, în practică, vom calcula întotdeauna valorile noastre din datele eșantionului, vom folosi de acum înainte simbolologia adecvată.

3.8 Scorurile Z și distribuția normală standardizată

Distribuția normală standardizată este una a cărei medie = 0, abaterea standard = 1 și aria totală sub curbă = 1. Distribuția normală standard arată ca cea prezentată în figură 3.4 .

Pe abscisă, în loc de ofx, avem o transformare ofx numită scor standard; Z. Z este derivat din x prin următoarele:

$$Z = \frac{x - m}{\sigma}$$

Astfel, scorul Z vă spune cu adevărat câte abateri standard de la medie este un anumit scor x.

Orice distribuție a unei variabile normale poate fi transformată într-o distribuție a lui Z luând fiecare valoare x, scăzând din ea media lui x (adică, m) și împărțind această abatere a lui x de la media ei, la abaterea standard. Să ne uităm din nou la distribuția IQ în Figura 3.5 .

Astfel, un scor IQ de 131 este echivalent cu un scor Z de 1,96 (adică este 1,96, sau aproape 2, abateri standard peste IQ-ul mediu).

$$Z = \frac{131 - 100}{16} \qquad 1{:}96$$

Unul dintre lucrurile frumoase despre distribuția Z este că probabilitatea ca o valoare să se afle oriunde între două puncte este egală cu aria de sub

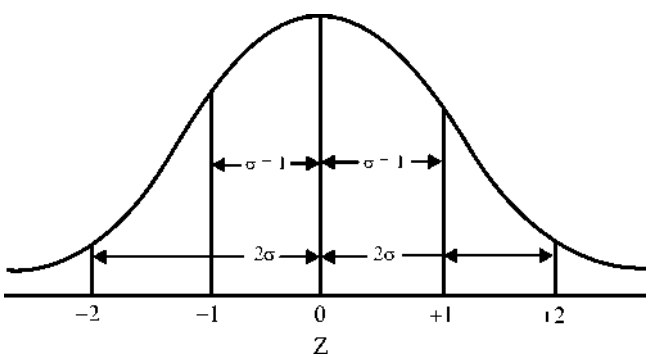


Figure 3.4 Standard normal distribution

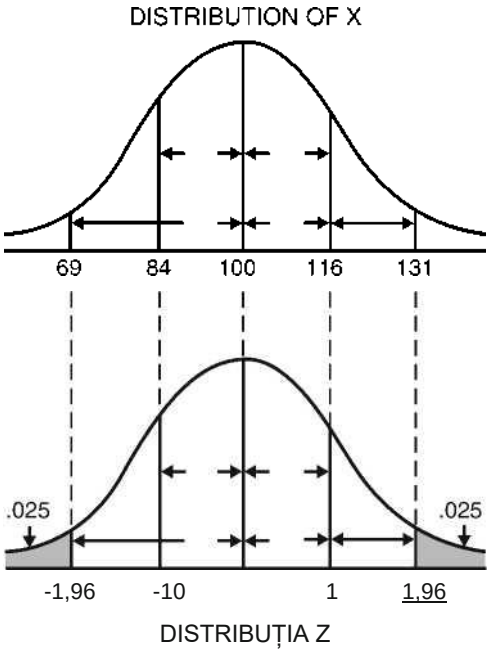


Figura 3.5 Distribuția scorurilor Z

curba dintre cele două puncte. (Acceptați acest lucru prin credință.) Se întâmplă ca aria din stânga lui 1,96 să corespundă unei probabilități de .025, sau 2,5 % din curba totală. Deoarece curba este simetrică, probabilitatea ca Z să fie la stânga lui —1,96 este de asemenea .025. Invocând legea probabilității aditive (Secțiunea 2.2), probabilitatea ca un Z să fie *la* stânga lui — 1,96 *sau* la stânga lui +1,96 este .025 + .025 = .05. Transformând înapoi la x, putem spune că probabilitatea ca cineva să aibă un IQ în afara de 1,96 deviații standard de la medie (adică peste 131 sau sub 69) este .05, sau doar 5% din populație are valori atât de extreme. . (De obicei, valoarea Z de 1,96 este rotunjită la ± 2 abateri standard de la medie ca corespunzătoare punctelor limită dincolo de care se află 5 % din curbă, dar valoarea exactă este 1,96.)

O utilizare foarte importantă a lui Z derivă din faptul că putem converti și o medie eșantion (mai degrabă decât o singură valoare individuală) într-un scor Z.

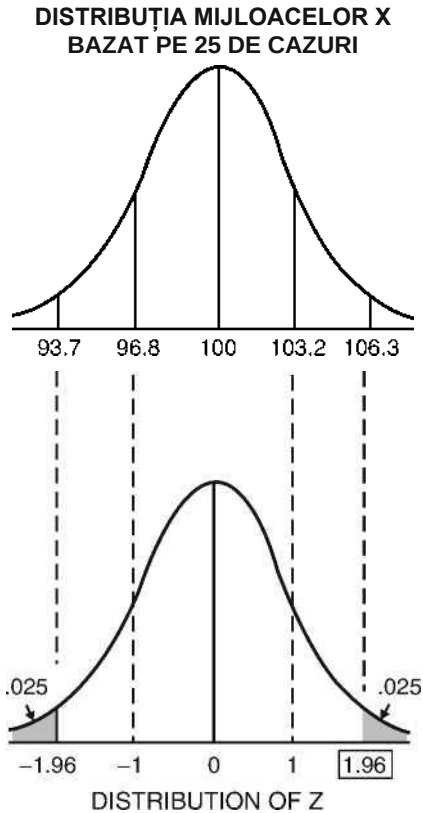


Figura 3.6 Distribuția mediilor pe baza eșantioanelor de mărimea 25

$$Z = \frac{x - m}{\sigma *}$$

Numătorul este acum distanța dintre media eșantionului de media populației, iar numitorul este abaterea standard a distribuției mediilor, care este eroarea standard a mediei. Acest lucru este ilustrat în figura 3.6 , unde luăm în considerare mijloace bazate pe 25 de cazuri fiecare. Se este $16 / \sqrt{25}$ ffi = 16 / 5 = 3 . 2.

Acum putem vedea că o medie a eșantionului de 106,3 corespunde unui scor Z de 1,96.

$$Z = \frac{106.3 - 100}{32} \qquad 1.96$$

Putem spune acum că probabilitatea ca IQ-ul mediu al unui grup de 25 de persoane să fie mai mare decât 106,3 este .025. Probabilitatea ca o astfel de medie să fie mai mică de 93,7 este, de asemenea, de .025.

Scorul AZ poate fi calculat și pentru diferența dintre două medii.

$$Z = \frac{(X_A - X_B) - (m_A - m_B)}{\sigma_{X_A - X_B}}$$

Dar $m_A - m_B$ este în mod obișnuit ipotezat a fi 0, astfel încât formula devine

$$Z = \frac{X_A - X_B}{\sigma_{X_A - X_B}}$$

Puteți vedea că un scor Z în general este o distanță între o anumită valoare și media ei împărțită la o eroare standard corespunzătoare.

Acest lucru devine foarte util mai târziu când vorbim despre intervalele de încredere în secțiuni 3.10 , 3.11 , 3.12 , 3.13, 3.14 și 3.15 .

3.9 Statistica t

Să presupunem că suntem interesați de mediile eșantionului și dorim să calculăm un scor Z. Nu știm care este abaterea standard a populației, dar dacă eșantioanele noastre sunt foarte mari, putem obține o estimare bună a σ calculând abaterea standard, sd, din eșantionul nostru și apoi obținem eroarea standard ca de obicei: s . e . = s . d ./ pñ. Dar adesea eșantionul nostru nu este suficient de mare. Putem obține în continuare un scor standardizat calculând o valoare numită student ' st:

$$t = \frac{x - m}{s.e._x}$$

Arată exact ca Z; singura diferență este că o calculăm din probă și este o probă mică.

Putem obține probabilitatea de a obține anumite valori t în mod similar modului în care am obținut probabilitățile de valori Z - dintr-un tabel adecvat. Dar se întâmplă că, în timp ce distribuția t arată ca o distribuție Z normală , este doar puțin diferită, dând astfel probabilități ușor diferite . De fapt, există multe distribuții t (nu doar una, ca pentru Z). Există o distribuție t diferită pentru fiecare dimensiune diferită a eșantionului. (Vor fi explicate mai multe despre acest lucru în secțiunea 3.10 .)

În exemplul nostru, unde avem o medie bazată pe 25 de cazuri, am avea nevoie de o valoare de 2,06 pentru a corespunde unei probabilități de .025 (în loc de 1,96 pentru distribuția Z). Traducând acest lucru înapoi la scara eșantionului înseamnă, dacă eroarea noastră standard ar fi 3,2, atunci probabilitatea ar fi de 0,025 ca să obținem o medie a eșantionului de 106,6 (care este 100 + 2,06 ori 3,2) mai degrabă decât 106,3 (care este 100 + 1,96 ori 3,2) ca în distribuția Z. Acest lucru poate părea o slăbiciune, deoarece diferențele sunt atât de mici. De fapt, pe măsură ce dimensiunea eșantionului se apropie de infinit, distribuția t devine exact ca distribuția Z, dar diferențele dintre Z și t devin mai mari pe măsură ce dimensiunea eșantionului devine mai mică și este întotdeauna sigur să utilizați distribuția t. De exemplu, pentru o medie bazată pe cinci cazuri, valoarea t ar fi 2,78 în loc de Z de 1,96. Unele valori t sunt prezentate în Anexa 1 . Tabelele mai detaliate sunt în cărțile de statistică standard.

3.10 Valori eșantioane și valori ale populației revizuite

Toate acestea mergând înainte și înapoi între valorile eșantionului și valorile populației pot fi confuze. Iată punctele de reținut:

1. Suntem întotdeauna interesați de estimarea valorilor populației din eșantioane.
2. În unele dintre formule și termeni, folosim valorile populației ca și cum am ști care sunt cu adevărat valorile populației. Desigur, nu cunoaștem valorile efective ale populației, dar dacă avem eșantioane foarte mari, le putem estima destul de bine din datele eșantionului nostru.
3. În scopuri practice, în general, vom folosi și ne vom referi la tehnici - adecvate pentru mostrele mici, deoarece acestea sunt mai frecvente și mai sigure (adică, nu strica chiar dacă avem mostre mari).

3.11 O chestiune de încredere

Un interval de încredere stabilește un interval și specifică probabilitatea ca acest interval să cuprindă media reală a populației. De exemplu, un interval de încredere de 95 % (aproximativ) este stabilit prin luarea mediei eșantionului, x̄, plus sau minus două erori standard ale mediei.

Interval de încredere de 95 %:

$$X \pm 2\,S\!:\!e\!:\! = x \pm 2\,\frac{s\!:\!d\!:\!}{n} =$$

Astfel, dacă am luat un eșantion aleatoriu de 64 de solicitanți la Colegiul de Medicină Albert Einstein și am constatat că IQ-ul lor mediu este de 125, să zicem, (o cifră fictivă), am dori să stabilim un interval de încredere de 95 % pentru a deduce ce adevărata medie a populației de solicitanți este într-adevăr. Intervalul de încredere de 95 % este intervalul cuprins între 125–2 se și 125 + 2 se.

> *„Suntem 95% încrezători că adevăratul IQ mediu al* candidaților la facultatea de medicină Einstein se află în 125 ± 2 se.”

În sensul acestui exemplu, presupuneți că abaterea standard este 16. (Acesta nu este o ipoteză deosebit de bună, deoarece variabilitatea IQ-ului solicitanților de la facultatea de medicină este considerabil mai mică decât variabilitatea IQ-ului în populație în general.) În conformitate cu această ipoteză, ajungem la următorul interval:

$$125 + \frac{2(16)}{\sqrt{64}} = 125 \pm \frac{2(16)}{8} = 125 \pm 4 = 121 - 129$$

Declarația noastră acum este următoarea: „Probabilitatea este de aproximativ .95 ca adevăratul IQ mediu al solicitanților de la Einstein Medical School să se încadreze în intervalul 121-129.” (O interpretare mai riguroasă a acestui lucru este dată în secțiunea 3.11 .)

interval de încredere de 99 % este aproximativ media eșantionului ± 3 se În exemplul nostru, acest interval ar fi

$$125 \pm 3 \frac{'(16)^{\wedge}}{. \sqrt{64.}} = 125 \pm 6 = 119 - 131$$

Atunci am putea spune: „Probabilitatea este de aproximativ .99 ca adevăratul IQ mediu al solicitanților de la Einstein Medical School să se încadreze în intervalul 119-131."

„Aproximativ" se datorează faptului că pentru a obține o probabilitate de .95, nu înmulțiți se cu 2 exact așa cum am făcut aici; l-am rotunjit pentru comoditate . Factorul de multiplicare exact depinde de cât de mare este eșantionul. Dacă eșantionul este foarte mare, mai mare de 100, am înmulți se cu 1,96 pentru intervale de încredere de 95 % și cu 2,58 pentru intervale de încredere de 99 %. Dacă eșantionul este mai mic, ar trebui să căutăm multiplicatorul în tabelele cu valorile t, care apar în multe texte. Aceste valori t sunt diferite pentru diferite „grade de libertate", explicate în secțiune 3.13 , care sunt legate de dimensiunile eșantionului. Unele valori t sunt prezentate în Anexa 1 . (De asemenea, consultați Secțiunea 3.9 pentru semnificația statisticilor t.)

Rețineți că pentru o anumită dimensiune a eșantionului, schimbăm gradul de certitudine cu dimensiunea intervalului. Putem fi mai siguri că adevărata noastră medie se află într-un interval mai larg, dar dacă dorim să stabilim intervalul mai precis, suntem mai puțin siguri despre ea (Figura 3.7). Pentru a obține mai multă precizie și pentru a menține o probabilitate ridicată de a fi corect în estimarea intervalului, este necesar să se mărească dimensiunea eșantionului. Principalul punct aici este că atunci când raportați o medie a eșantionului ca o estimare a unei medii populației, este cel mai de dorit să raportați limitele de încredere.

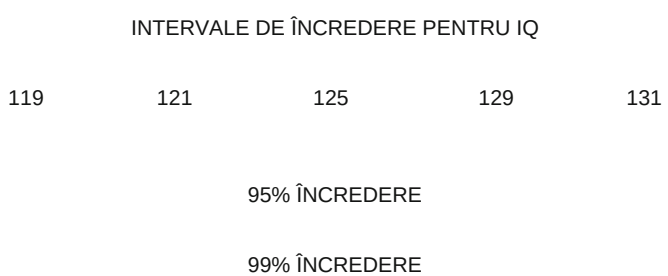


Figura 3.7 Intervale de încredere

3.12 de **încredere și intervale** de încredere

Limitele de încredere sunt granițele exterioare pe care le calculăm și despre care putem spune: suntem 95 % încrezători că aceste limite sau limite includ adevărata medie a populației. Intervalul dintre aceste limite se numește interval de încredere. Dacă ar fi să luăm un număr mare de eșantioane din populație și să calculăm limitele de încredere de 95 % pentru fiecare dintre ele, 95 % din intervalele limitate de aceste limite ar conține media reală a populației. Cu toate acestea, 5 % nu l-ar conține. Desigur, în viața reală luăm doar un eșantion și construim intervale de încredere din acesta. Nu putem fi niciodată siguri dacă intervalul calculat din eșantionul nostru particular este unul dintre cele 5 % dintre astfel de intervale care nu conțin media populației. Cel mai mult putem spune că suntem 95% încrezători că îl conține. După cum puteți vedea, nu știm niciodată nimic sigur.

Dacă din populația de interes au fost extrase un număr infinit de eșantioane aleatoare independente (cu înlocuire), atunci 95 % din intervalele de încredere calculate din eșantioane (cu media $\bar{x}$ și eroarea standard se) vor cuprinde adevărata medie a populației m.

Figura 3.8 ilustrează conceptele de mai sus.

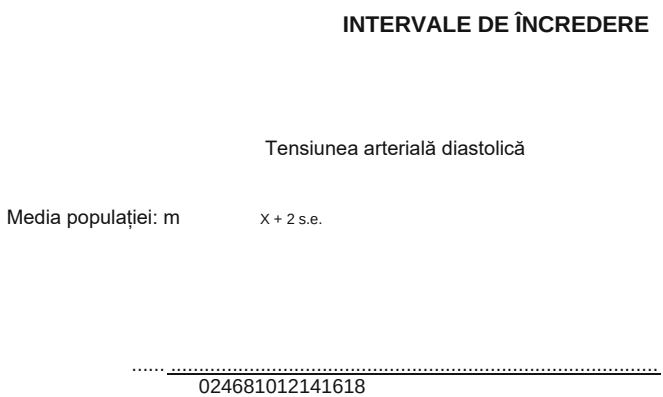


Figura 3.8 Intervale de încredere

3.13 Grade de libertate

Valorile t pe care le folosim ca multiplicator al erorii standard pentru a construi intervale de încredere depind de ceva numit grade de libertate (df), care sunt legate de dimensiunea eșantionului. Când avem un eșantion, pentru a găsi valoarea t adecvată pentru a calcula limitele de încredere, introducem tabelele cu n–1 grade de libertate, unde n este dimensiunea eșantionului. O modalitate intuitivă de a înțelege conceptul de df este să luăm în considerare că, dacă calculăm media unui eșantion de, să zicem, trei valori, am avea „libertatea" de a varia două dintre ele în orice fel ne-ar plăcea după ce știm care este media. , dar al treilea trebuie fixat pentru a ajunge la media dată. Deci avem doar 2 „grade de libertate". De exemplu, dacă știm că media a trei valori este 7, putem avea următoarele seturi de date:

Valoarea 1:	7	-50
Valoarea 2:	7	+18
Valoarea 3:	7	+53
Suma =	21	21
Media =	X = 7	x = 7

În fiecare caz, dacă cunoaștem valorile 1 și 2, atunci se determină valoarea 3 deoarece suma acestor valori trebuie să fie 21 pentru ca media să fie 7. Am „pierdut" un grad de libertate în calcularea mediei.

3.14 de **încredere pentru proporții**

O proporție poate fi considerată o variabilă continuă. De exemplu, în studiul anticoagulant descris în secțiunea 3.1 , proporția femeilor din grupul de control (tratat cu placebo) care au supraviețuit unui atac de cord a fost de 89/129 = 0,69. O proporție poate lua valori de-a lungul continuumului între 0 și 1. Putem construi un interval de încredere în jurul unei proporții într-un mod similar cu construirea intervalelor de încredere în jurul mediilor. Limitele de încredere de 95 % pentru o proporție sunt $p \pm 1{,}96\ se_p$, unde se_p este eroarea standard a unei proporții.

Pentru a calcula eroarea standard a unei proporții, trebuie mai întâi să calculăm abaterea standard a unei proporții și să o împărțim la rădăcina pătrată a

lui n. Ne definim simbolologia:

$$s = \text{abaterea standard a unei proporții} = \sqrt{pq}$$

$$P = \text{sample proportion} = \frac{\text{numărul de supraviețuitori din grupul de control}}{\text{numărul total de femei din grupul de control}}$$

$$q = 1 - P = \frac{\text{numărul morților în grupul de control}}{\text{numărul total de femei din grupul de control}}$$

$$\text{s.e.}P = \sqrt{\frac{pq}{n}}$$

În exemplul nostru de femei care au supraviețuit unui atac de cord din grupul de control, intervalul de încredere de 95 % este

$$.69 \pm 1.96 \times \sqrt{\frac{.69 \times .31}{129}} = .69 \pm .08$$

Și putem afirma că suntem 95% încrezători că proporția populației de femei netratate care supraviețuiesc unui atac de cord este între .61 și .77 sau 61% și 77%. (Rețineți că aceasta se referă la populația din care a fost extras eșantionul nostru. Nu putem generaliza acest lucru la toate femeile care au un atac de cord.)

Pentru limitele de încredere de 99 %, am înmulți eroarea standard a unei proporții cu 2,58, pentru a obține intervalul de la .59 la .80. Multiplicatorul este valoarea Z care corespunde cu 0,95 pentru limitele de încredere de 95 % sau cu probabilitatea 0,99 pentru limitele de încredere de 99 %.

3.15 de **încredere în jurul diferenței dintre două mijloace**

Putem construi intervale de încredere în jurul unei diferențe dintre medii într-un mod similar cu care am construit intervale de încredere în jurul unei singure medii. Limitele de încredere de 95 % în jurul diferenței dintre medii sunt date de

$$(x - y) \pm f_{,.95)}(s \cdot e \cdot x - y)$$

Cu cuvinte, aceasta este diferența dintre cele două medii de probă, plus sau minus o valoare t adecvată, ori eroarea standard a diferenței; df este gradele de libertate și .95 spune că căutăm valoarea t care se referă la acele grade de libertate și la probabilitatea .95. Gradele de libertate atunci când ne uităm la două eșantioane sunt $n_x + n_y - 2$. Acest lucru se datorează faptului că am pierdut un grad de libertate pentru fiecare dintre cele două mijloace pe care le-am calculat, deci gradele noastre totale de libertate sunt $(n_x - 1) + (n_y - 1) = n_x + n_y — 2$.

De exemplu, luați în considerare că avem un eșantion de 25 de studenți la medicină și 25 de bărbați. IQ-urile medii pentru fiecare probă sunt

$$x_{females} \quad 130, \quad x_{males} \quad 126, \quad S_{pooled} = 12, \quad \wedge f = 48$$

Intervalul de încredere de 95 % pentru diferența medie între bărbați și femei se calculează după cum urmează:

tabelele t , aflăm că valoarea t pentru df = 48 este 2,01:

$$x_{femele} \quad {}^{x}{}_{masculi} \; I\,2\,.\,01 \times$$

$$(130 — 126) \pm 2\,.\,01\sqrt{}12(1/25 + 1\,/\,25) = 4 \pm 6\,.\,8$$

Intervalul este atunci de —2,8 până la 10,8 și suntem 95% siguri că include adevărata diferență medie dintre bărbați și femei. Acest interval include 0 diferență, așa că ar trebui să concluzionam că diferența de IQ între bărbați și femei poate fi zero.

3.16 Comparații între două grupuri

O problemă cea mai frecventă care apare este necesitatea de a compara două grupuri pe o anumită dimensiune. Am putea dori să stabilim, de exemplu, dacă (1) administrarea unui anumit medicament scade tensiunea arterială sau (2) medicamentul A este mai eficient decât medicamentul B în scăderea nivelului zahărului din sânge sau (3) învățarea copiilor de clasa întâi să citească prin metoda I produce scoruri mai mari la sfârșitul anului de citire decât învățarea citirii prin metoda II.

3.17 Testul Z pentru compararea a două proporții

Ca exemplu, reproducem aici tabelul din Secțiune 3.1 care prezintă date dintr-un studiu privind terapia anticoagulantă.

	Frecvențele observate		
	Controla	Tratat	
A trăit	89	223	312
A murit	40	39	79
Total	129	262	391

Dacă dorim să testăm dacă proporția de femei care au supraviețuit unui atac de cord în grupul tratat diferă de proporția care au supraviețuit în grupul de control, am stabilit ipoteza noastră nulă ca

H_O: $P_1 = P_2$ sau $P_1 - P_2 = 0$; P_1 = proporția supraviețuitoare în populația tratată

P_2 = proporția de supraviețuire în populația martor

H_A: $P_1 - P_2 = 0$ (diferența nu este egală cu 0)

Noi calculăm

$$Z = \frac{P_1 \quad P_2}{S : e_{P_1 - P_2}}$$

$p^{i} = \frac{223}{262} = : 85$, $q_1 = 1 - p_1 = : 15$, $n_1 - 262$

$p^2 = \frac{89}{129} = : 69$, $q_2 = 1 - p_2 = : 31$, $\Pi_2 = 129$

Astfel, numărătorul lui Z = .85-.69 = .16.

Numitorul =

standard a diferenței dintre două proporții =

$$s \cdot e \frac{3}{8} \mathbf{1} \cdot p\hat{} =$$

unde *p* și *q* sunt estimări combinate bazate atât pe cele tratate, cât și pe cele de control
date de grup. O calculăm după cum urmează:

$$\hat{P} \quad \frac{n_1 p_1 + n_2 P_2}{n_1 + n_2}$$

numărul de supraviețuitori în tratați + control
numărul total de pacienți în tratați + control

$$= \frac{262(.85) + 129(.69)}{262 + 129} = \frac{223 + 89}{391} = .80$$

$$q = 1 - p = 1 - .80 = .20$$

$$s \cdot e\frac{1}{8} - \kappa) = \sqrt{(.80)(.20)\left(\frac{1}{129} + \frac{1}{262}\right)} = .043$$

$$Z = \frac{.85 - .69}{.043} \quad 3.72$$

Trebuie să ne uităm acum pentru a vedea dacă această valoare a lui Z depășește valoarea critică. Valoarea critică este valoarea minimă a statisticilor de test pe care trebuie să o obținem pentru a respinge ipoteza nulă la un anumit nivel de semnificație.

Valoarea *critică a lui Z* pe care trebuie să o respingem H_0 la nivelul de semnificație 0,05 este 1,96. Valoarea pe care am obținut-o este 3,74. Acesta este în mod clar un Z suficient de mare pentru a respinge H_0 cel puțin la nivelul .01. Valoarea critică pentru ca Z să respingă H_0 la nivelul 0,01 este 2,58.

Rețineți că am ajuns la aceeași concluzie folosind testul chi-pătrat din secțiune 3.1 . De fapt $Z^2 = \chi^2 = (3,74)^2 = 13,99$ și chi-pătratul necorectat pe care l-am calculat a fost 13,94 (diferența se datorează erorilor de rotunjire). Desigur, valorile critice ale lui χ^2 și Z trebuie căutate în tabelele corespunzătoare. Unele valori apar în Anexa 1 .

3.18 t-Test pentru diferența dintre mijloacele a două grupuri independente: principii

Când am vrut să comparăm două grupuri pe o anumită măsură care a fost o variabilă discretă sau categorică, cum ar fi mortalitatea în două grupuri, am folosit testul chi-pătrat, descris în secțiunea 3.1 . Sau am putea folosi un test între proporții așa cum este descris în secțiune 3.17 . Acum discutăm o metodă de comparare a două grupuri atunci când măsura interesului este o variabilă continuă.

Să luăm ca exemplu comparația vârstelor la prima sarcină a două grupuri de femei: cele care sunt avocate și cele care sunt asistente juridice. Un astfel de studiu ar putea fi de interes sociologic sau ar putea fi de interes pentru firmele de avocatură sau poate pentru o companie de alimente pentru copii care încearcă să-și concentreze strategia de publicitate mai eficient.

Presupunând că am luat eșantioane adecvate din fiecare grup, acum avem două seturi de valori: vârstele avocaților (grupa A) și vârstele parajuriștilor (grupa B), și avem o vârstă medie pentru fiecare eșantion. Am stabilit ipoteza noastră nulă după cum urmează:

H_O: „Vârsta medie a populației de avocați de la care avem eșantionul extras A este același cu vârsta medie a populației de parajuriști din care am extras eșantionul B.”

Ipoteza noastră alternativă este

H_A: „Vârstele medii ale celor două populații pe care le-am eșantionat sunt diferit."

În esență, deci, am extras eșantioane pe baza cărora vom face inferențe despre populațiile din care provin. Suntem supuși acelorași tipuri de erori de tip I și de tip II pe care le-am discutat anterior.

Abordarea generală este următoarea. Știm că există o variabilitate a scorurilor în grupul A în jurul mediei pentru grupul A și în cadrul grupului B în jurul mediei pentru grupul B, pur și simplu pentru că chiar și într-o anumită populație, oamenii variază. Ceea ce dorim să aflăm este dacă variabilitatea dintre cele două medii ale eșantionului în jurul mediei mari a tuturor scorurilor este mai mare decât variabilitatea vârstelor din cadrul grupurilor în jurul propriilor medii.

Dacă există la fel de multă variabilitate în cadrul grupurilor ca și între grupuri, atunci probabil că aceștia provin din aceeași populație.

Testul potrivit aici este testul t. Calculăm o valoare cunoscută sub numele de t, care este egală cu diferența dintre cele două medii de eșantion împărțită la o eroare standard corespunzătoare. Eroarea standard adecvată se numește eroarea standard a diferenței dintre două medii și se scrie ca

$$s : e : X_1 — X \mathbf{2}$$

Distribuția lui t a fost tabelată, iar din tabele, putem obține probabilitatea de a obține o valoare a lui t la fel de mare ca cea pe care am obținut-o de fapt, în ipoteza că ipoteza noastră nulă (de nicio diferență între medii) este adevărată. Dacă această probabilitate este mică (adică, dacă este puțin probabil ca doar din întâmplare să obținem o valoare a t atât de mare dacă ipoteza nulă ar fi adevărată), am respinge ipoteza nulă și am accepta ipoteza alternativă că există într-adevăr o diferență. între mediile populațiilor din care am extras cele două eşantioane.

3.19 Cum se face un t -test: un exemplu

Deși testele t pot fi efectuate cu ușurință pe computere personale, un exemplu de calcule și interpretare este dat mai jos. Acest test statistic este efectuat pentru a compara mediile a două grupuri în ipoteza că ambele eșantioane sunt aleatoare, independente și provin din populații distribuite normal cu variante necunoscute, dar egale.

Ipoteza nulă : $m_A = m_B$, sau echivalentul : $m_A — m_B = 0$

Ipoteza alternativă : $m_A = m_B$ *sau* echivalentul : $m_A — m_B = 0$

[Notă: Când ipoteza alternativă nu specifică direcția diferenței (prin afirmarea, de exemplu, că m_A este mai mare decât m_B), ci pur și simplu spune că diferența *nu este egală cu 0,* se numește test cu două cozi. Când este specificată direcția diferenței, se numește test cu o singură coadă. Mai multe despre acest subiect apar în Secțiunea 5.4 .]

$$= \frac{(X_A — X_B)}{s_{X_A — X_B}}$$

Vârstele probei A			Vârstele probei B		
x_i	$(x_i - x_A)$	$(x_i - x_A)^2$	x_i	$(x_i - x_B)$	$(x_i - x_B)^2$
28	- 3	9	24	2.4	5,76
30	- 1	1	25	3.4	11.56
27	- 4	16	20	- 1.6	2,56
32	1	1	18	- 3.6	12.96
34	3	9	21	- 0,6	0,36
36	5	25	Σ ■ 108	Σ ■ 0	Σ ■ 33.20
30	- 1	1			
Σ ■ 217	Σ ¼ 0	Σ ■ 62			

$$Media_A = \frac{\Sigma x_i}{n} = \frac{217}{7} = 31 \; ; \; Media_B = \frac{\Sigma x_i}{n} = \frac{108}{5} = 21.6$$

(Indicele *i* se referă la scorul *i* și este o convenție folosită pentru a indica faptul că însumăm toate scorurile.)

Numătorul lui t este diferența dintre cele două medii:

$$31 - 21 : 6 = 9 : 4$$

Pentru a obține numitorul lui t, trebuie să calculăm eroarea standard a diferenței dintre medii, ceea ce facem după cum urmează:

Mai întâi obținem estimarea cumulativă a abaterii standard. Calculam:

$$Sp = \frac{\delta\text{-}i - X_A)^2 + \Sigma(x_{¿} - \text{-}_B)^2}{n_A + n_в - 2} = \frac{62 + 33:20}{7 + 5 - 2}$$

$$¼V^z \; \frac{95:20}{10} \quad 9:52 \blacksquare 3:09$$

$$sx. \; {}_{AB} - X\beta \blacksquare sp\lfloor I - + I \blacksquare 3 \cdot 09\sqrt{\frac{1}{nA} + \frac{1}{nB}} \blacksquare 3.09\sqrt{.3428} \blacksquare 3.09 \times .5854 \blacksquare 1.81$$

$$t \blacksquare \frac{X_A \quad \hat{x}\hat{B}}{s_{XA} - X_B} \quad \frac{9:4}{1:81} \blacksquare 5.19$$

Acest t este semnificativ la nivelul .001, ceea ce înseamnă că ați obține o valoare a t la fel de mare ca aceasta sau mai mare doar o dată din o mie din

întâmplare dacă ipoteza nulă ar fi adevărată. Așa că respingem ipoteza nulă a lipsei de diferență, acceptăm ipoteza alternativă și concluzionăm că avocații sunt mai în vârstă la prima sarcină decât parajuriștii.

3.20 Testul t de perechi potrivite

Dacă aveți o situație în care scorurile dintr-un grup se corelează cu scorurile din celălalt grup, nu puteți utiliza testul t obișnuit, deoarece presupune că cele două grupuri sunt independente. Această situație apare atunci când luați două măsuri asupra aceluiași individ. De exemplu, să presupunem că grupa A reprezintă scorurile de citire ale unui grup de copii luate la momentul 1. Acești copii au primit apoi instruire specială în citire pe o perioadă de 6 luni și rezultatele lor în lectură sunt din nou măsurate pentru a vedea dacă au obținut vreun câștig la timp 2. Într-o astfel de situație, ați folosi un test t de pereche potrivită.

Copil	O Scorurile inițiale de citire ale copiilor	B Scoruri ale acelorași copii după 6 luni de antrenament	d = BA	$d - d$	$(d - d)^2$
(1)	60	62	2	1.4	1,96
(2)	50	54	4	3.4	11.56
(3)	70	70	0	— 0,6	0,36
(4)	80	78	- 2	— 2.6	6,76
(5)	75	74	- 1	— 1.6	2,56
		Sumă	3	0	23.20
Diferența medie $= d = 3 / 5 = 0.60$					

Ipoteza nulă: diferența medie = 0.

Ipoteza alternativă: diferența medie este mai mare decât 0.

$$= d \; . \; s_d \; ; \; s_d \quad \frac{s}{\sqrt{n}}$$

$$s = \frac{.1 \; \Sigma(d - d)2}{'23\ 20 \; :} \quad 75{:}8 = 2.41$$

$$s_d \quad \frac{2{:}41}{\sqrt{5}} \quad \frac{2{:}41}{2{:}23} \quad 1{:}08$$

$$t = \frac{:60}{1{:}08}$$

Acest t nu este semnificativ, ceea ce înseamnă că nu respingem ipoteza nulă și concluzionăm că diferența medie a scorurilor la citire ar putea fi zero; adică este posibil ca programul de lectură de 6 luni să nu fie eficient. (Sau poate că studiul nu a fost suficient de mare pentru a detecta o diferență și am comis o eroare de tip II.)

Atunci când diferența reală dintre perechile potrivite nu este în sine un număr semnificativ, dar cercetătorul poate clasifica scorurile diferențelor (ca fiind mai mari sau mai mici pentru perechi date), testul potrivit este testul Wilcoxon pentru sumele de rang pentru perechi potrivite. Acesta este cunoscut ca un test neparametric și, împreună cu alte astfel de teste, este descris cu o claritate deosebită în cartea clasică a lui Sidney Siegel, Statistics nonparametrice pentru științe comportamentale și în Statistici nonparametrice pentru non-statisticieni (enumerate în lecturi sugerate).

3.21 Când să nu faceți o mulțime de teste t: problema testelor multiple de semnificație

Un test t este utilizat pentru a compara mediile a două grupuri. Când există trei sau mai multe medii de grup de comparat, testul t nu este adecvat. Pentru a înțelege de ce, trebuie să ne invocăm cunoștințele despre combinarea probabilităților din Secțiunea 2.2 .

Să presupunem că testați efectele a trei tratamente diferite pentru hipertensiune arterială. Pacienții dintr-un grup A primesc un singur medicament, un diuretic; pacienții din grupa B primesc un alt medicament, un beta-blocant; iar pacientii din grupul C primesc o pastila placebo. Vrei să știi dacă oricare dintre medicamente este mai bun decât placebo în scăderea tensiunii arteriale și dacă cele două medicamente sunt diferite unul de celălalt în ceea ce privește efectul lor de scădere a tensiunii arteriale.

Există trei comparații care pot fi făcute: grupul A față de grupul C (pentru a vedea dacă diureticul este mai bun decât placebo), grupul B față de grupul C (pentru a vedea dacă beta-blocantul este mai bun decât placebo) și grupul A față de grupa B (pentru a vedea care dintre cele două medicamente active are mai mult efect). Ne-am stabilit nivelul de semnificație la .05, adică suntem dispuși să greșim respingând ipoteza nulă a lipsei de diferență între două medii, cu o probabilitate de .05 sau mai puțin (adică, probabilitatea noastră de a face o eroare de tip I trebuie să nu fie mai mare de .05). Luați în considerare următoarele:

Comparație	Probabilitatea erorii de tip I	Probabilitatea de *a nu* face o eroare de tip I = 1 – *P* (eroare de tip I)
1. A versus C	.05	1-.05 = .95
2. B versus C	.05	1-.05 = .95
3. A versus B	.05	1-.05 = .95

Probabilitatea de a nu face o eroare de tip I la prima comparație și de a nu o face în a doua comparație și de a nu o face în a treia comparație = .95 × .95 × .95 = .86. Ne uităm aici la apariția *în comun* a trei evenimente (cele trei moduri de a *nu* comite o eroare de tip I) și combinăm aceste probabilități prin înmulțirea probabilităților individuale. (Nu uitați, când vedem „și” în contextul combinării probabilităților, înmulțim, când vedem „sau” adăugăm.) Așadar, acum, știm că probabilitatea globală de a *nu* comite o eroare de tip I în oricare dintre trei comparații posibile este .86. Prin urmare, probabilitatea de a comite o astfel de eroare este 1 - probabilitatea de a nu o comite, sau 1-.86 = .14. Astfel, probabilitatea globală a unei erori de tip I ar fi considerabil mai mare decât .05 pe care l-am specificat ca fiind de dorit. De fapt,

numerele sunt puțin diferite deoarece cele trei comparații nu sunt evenimente

independente, deoarece aceleași grupuri sunt utilizate în mai multe comparații, astfel încât combinarea probabilităților în această situație nu ar implica regula simplă a înmulțirii pentru apariția în comun a evenimentelor independente. Cu toate acestea, este suficient de aproape pentru a ilustra punctul că efectuarea de comparații multiple în același experiment are ca rezultat un nivel de semnificație destul de diferit (.14 în acest exemplu) față de cel pe care l-am ales (.05). Când există mai mult de trei grupuri de comparat, situația se înrăutățește.

3.22 Analiza variației: comparație între Mai multe Grupuri

Tehnica adecvată pentru analiza variabilelor continue atunci când există trei sau mai multe grupuri de comparat este analiza variației, denumită în mod obișnuit ANOVA. Un exemplu ar putea fi compararea efectelor de reducere a tensiunii arteriale ale celor trei medicamente.

3.23 Principii care stau la baza analizei variației

Principiile implicate în analiza variației sunt aceleași cu cele din testul t. În ipoteza nulă, am avea următoarea situație : ar exista o populație mare și dacă am alege eșantioane de o anumită dimensiune din acea populație, am avea o grămadă de medii ale eșantionului care ar varia în funcție de șansă în jurul mediei mari. a întregii populaţii. Dacă se dovedește că variază în jurul valorii mari mai mult decât ne-am aștepta doar din întâmplare, atunci poate că altceva decât șansa funcționează. Poate că nu toți provin din aceeași populație . Poate că ceva distinge grupurile pe care le-am ales. Am respinge apoi ipoteza nulă conform căreia toate mijloacele sunt egale și am concluziona că mijloacele sunt diferite unele de altele prin mai mult decât întâmplător. În esență, dorim să știm dacă variabilitatea tuturor mediilor de grup este substanțial mai mare decât variabilitatea în cadrul fiecăruia dintre grupuri în jurul propriei medii.

Calculăm o cantitate cunoscută sub numele de varianță între grupuri, care este variabilitatea mediilor grupului în jurul mediei mari a tuturor datelor. Calculăm o

altă cantitate numită varianță în interiorul grupurilor, care este variabilitatea scorurilor din cadrul fiecărui grup în jurul propriei sale medii. Una dintre ipotezele analizei variației este că amploarea variabilității indivizilor în cadrul grupurilor este aceeași pentru fiecare dintre grupuri, astfel încât putem pune în comun estimările variațiilor individului în cadrul grupului pentru a obține o estimare mai fiabilă a totalului în cadrul grupurilor. -varianta grupurilor. Dacă există atâta variabilitate a indivizilor în cadrul grupurilor cât există variabilitate a mediilor între grupuri, probabil că mediile provin din aceeași populație, ceea ce ar fi în concordanță cu ipoteza unei diferențe adevărate între medii, adică nu am putea respinge ipoteza nulă a lipsei de diferență între medii.

Raportul dintre variația dintre grupuri și varianța în interiorul grupurilor este cunoscut sub numele de raportul F. Valorile distribuției F apar în tabele în multe texte statistice, iar dacă valoarea obținută din experimentul nostru este mai mare decât valoarea critică care este tabelată, atunci putem respinge ipoteza lipsei de diferență.

Există diferite valori critice ale lui F, în funcție de câte grupuri sunt comparate și de câte scoruri există în fiecare grup. Pentru a citi tabelele lui F, trebuie să cunoaștem cele două valori ale gradelor de libertate (df). Df corespunzătoare variației între grupuri, care este numărătorul raportului F, este egal cu k–1, unde k este numărul de grupuri. Df-ul corespunzător numitorului raportului F, care este varianța în interiorul grupurilor , este egal cu $k \times (n-1)$, adică numărul de grupuri înmulțit cu numărul de scoruri din fiecare grup minus unu. De exemplu, dacă în experimentul nostru de hipertensiune există 100 de pacienți în fiecare dintre cele trei grupuri de medicamente, atunci gradele de libertate ale numărătorului ar fi 3-1 = 2, iar gradele de libertate numitorului ar fi $3 \times 99 = 297$. Un F raportul ar trebui să fie de cel puțin 3,02 pentru un nivel de semnificație de 0,05. Dacă ar fi comparate patru grupuri, atunci gradele de libertate ale numărătorului ar fi 3, iar valoarea critică a lui F ar trebui să fie 2,63. Dacă nu există un număr egal de indivizi în fiecare grup, atunci numitorul gradelor de libertate este $(n_1 - 1) + (n_2 - 1) + (n_3 - 1)$.

Nu vom prezenta aici calculele efective necesare pentru a face un test F pentru ca in zilele noastre acestea se fac rareori manual. Există o mare
numărul de programe disponibile pentru computerele personale care pot efectua teste F, teste t și majoritatea altor analize statistice. Cu toate acestea, mai jos este prezentat tipul de rezultate la care se poate aștepta de la aceste programe. Sunt prezentate date rezumative din studiul TAIM (Trial of Antihypertensive Interventions and Management). Studiul TAIM a fost conceput pentru a evalua

efectul dietei și al medicamentelor, utilizate singure sau în combinație între ele, pentru a trata persoanele supraponderale cu hipertensiune ușoară (tensiune arterială crescută) [10,11].

Următorul tabel arată scăderea medie a tensiunii arteriale după 6 luni de tratament cu fiecare medicament, numărul de persoane din fiecare grup și abaterea standard a modificării tensiunii arteriale în fiecare grup.

Grup de droguri	n	Scădere medie (în unități de tensiune arterială diastolică după 6 luni de tratament)	Abaterea standard
A. Diuretic	261	12.1	7.9
B. Beta-blocant	264	13.5	8.2
C. Placebo	257	9.8	8.3

Următorul tabel rezultă dintr-o analiză a varianței datelor din acest studiu. Acesta trebuie interpretat astfel:

ANOVA					
Sursa de variație	Grade de libertate	Suma pătratelor	Pătrat mediu	Raportul F	P 2 > F
Între grupuri	2	1776,5	888,2	13.42	.0001
În cadrul grupurilor	779	5256,9	66.2		
	781				

Pătratul mediu este suma pătratelor împărțită la gradele de libertate. Pentru între grupuri, este variația mediei de grup în jurul mediei mari, în timp ce pentru grupurile în interior, este estimarea combinată a variației scorurilor individuale în jurul valorii medii ale grupului respectiv.

Pătratul mediu din interiorul grupurilor se mai numește și pătratul mediu al erorii. (Un punct important este că rădăcina pătrată a pătratului mediu al erorii este estimarea combinată a abaterii standard în cadrul grupurilor. În acest caz, este √66 : 2 = 8 : 14. Este aproximativ echivalentă cu abaterea standard medie . .) F este raportul dintre medii pătrate și medii; în acest exemplu, este 888,2/66,2 = 13,42.

Raportul F este semnificativ la nivelul .0001, deci putem respinge ipoteza nulă că toate mediile de grup sunt egale. Cu toate acestea, nu știm unde este diferența. Este grupul A diferit de grupul C, dar nu de grupul B? Nu ar trebui pur și simplu să facem posibile toate comparațiile pe perechi din cauza problemei comparațiilor multiple discutate mai sus. Dar există modalități de a gestiona această problemă. Una dintre ele este procedura Bonferroni, descrisă în secțiunea următoare.

3.24 Procedura Bonferroni: O abordare a realizarii Comparații multiple

Aceasta este o modalitate de a gestiona problema comparațiilor multiple. Procedura Bonferroni implică că, dacă, de exemplu, facem cinci comparații , probabilitatea ca niciuna dintre cele cinci valori p să nu scadă sub .05 este de cel puțin 1-(5 × .05) = .75 atunci când ipoteza nulă a mediei egale. este cu adevărat adevărat. Aceasta înseamnă că există o probabilitate de până la .25 ca cel puțin o valoare p să atingă nivelul de semnificație de .05 doar întâmplător, chiar dacă tratamentele nu diferă într-adevăr. Pentru a ocoli acest lucru, împărțim nivelul de semnificație general ales la numărul de comparații în două sensuri care trebuie făcute, considerăm că această valoare este nivelul de semnificație pentru orice comparație unică și respingem ipoteza nulă a lipsei de diferență numai dacă atinge acest lucru. nou nivel de semnificație.

De exemplu, dacă dorim un nivel general de semnificație de .05 și vom face trei comparații între medii, ar trebui să atingem nivelul .05/3 = .017 pentru a respinge ipoteza nulă și a concluziona că există o diferență între două mijloace. Este necesară o bună autodisciplină pentru a respecta această procedură și pentru a nu declara că diferența dintre două mijloace este puțin probabil să se datoreze întâmplării dacă comparația particulară are semnificație la $p = .03$, să zicem, în loc de .017.

Procedura Bonferroni nu necesită un test F prealabil. Să aplicăm procedura Bonferroni datelor noastre.

Mai întâi comparăm fiecare dintre medicamente cu placebo. Se calculează t pentru diferența dintre mediile grupului A față de grupul C.

$$t = \frac{*_A \quad \text{XC}}{s.e._{X_A - X_C}}$$

$$s.e._{X_A - X_C} \quad s_p \quad n_A \, n_C$$

$$\frac{12{:}1 - 9{:}8}{8{:}14\sqrt{261 + 257}} = \frac{}{\cdot 715} - 3{:}22$$

$$p - {:}\ 0014$$

Rețineți că folosim 8.14 ca *s* pooled. Am obținut acest lucru din analiza variației ca estimare a abaterii standard comune. Gradele de libertate pentru a intra în tabelele t sunt 261 +257-2 — 516.

Se pare că probabilitatea de a obține o valoare t atât de mare din întâmplare este de numai .0014. Putem spune cu siguranță că diureticul reduce tensiunea arterială mai mult decât placebo. Același lucru este valabil și pentru comparația dintre beta-blocant și placebo. Acum să comparăm cele două medicamente, B versus A:

$$\frac{13{:}5 - 12.1}{8{:}14\sqrt{264 + 261}} \qquad \frac{1{:}4}{{:}711} - 1{:}97$$

Valoarea *p* corespunzătoare acestei valori t este .049. Ar putea fi tentant să declarăm o diferență semnificativă la nivelul .05, dar rețineți că procedura Bonferroni necesită să obținem o valoare *p* de .017 sau mai puțin pentru semnificație ajustată pentru comparații multiple. Valoarea critică a lui t corespunzătoare *lui p* — .017 este 2,39 și am ajuns doar la 1,97. Cu toate acestea, s-au pus sub semnul întrebării ajustarea de rutină pentru comparații multiple [12] pe motiv că astfel putem comite mai multe erori de tip II și pierdem efecte importante. În orice caz, nivelurile *p* ar trebui raportate, astfel încât cititorul informat să poată evalua dovezile.

3.25 Analiza varianței atunci când există două variabile independente: ANOVA cu doi factori

Exemplul de mai sus este denumit ANOVA unidirecțională deoarece puteți împărți toate scorurile într-un singur fel, în funcție de grupul de medicamente căruia i-au fost alocați pacienții. Grupul de droguri se numește „factor", iar acest factor are trei niveluri, ceea ce înseamnă că există trei categorii de medicamente. Totuși, poate exista un alt factor care clasifică indivizii și, în acest caz, am avea un ANOVA cu două sensuri sau cu doi factori. În experimentul pe care l-am folosit ca exemplu, pacienții au fost repartizați la unul dintre cele trei medicamente menționate mai sus, precum și la unul dintre cele trei regimuri de dietă - scădere în greutate, restricție de sodiu (sare) sau nicio schimbare față de dieta lor obișnuită, ceea ce este analog la o condiție de dietă placebo. Diagrama de mai jos ilustrează acest design cu doi factori și scăderea medie a tensiunii arteriale în fiecare grup, precum și numărul de cazuri din fiecare grup, care sunt prezentate în paranteză.

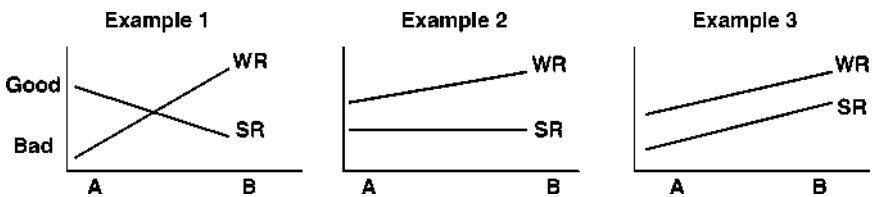
Acum suntem interesați să comparăm cele trei mijloace care reprezintă modificarea tensiunii arteriale în grupele de medicamente, cele trei mijloace care reprezintă schimbări în grupele de dietă și interacțiunea dintre medicament și dietă. Acum explicăm conceptul de interacțiune.

	Dieta			
Medicament	De obicei	Reducere în greutate	Restricție de sodiu	Total
Diuretic	10.2 (87)	14.5 (86)	11.6 (88)	12.1 (261)
Beta-blocant	12.8 (86)	15.2 (88)	12.6 (90)	13.5 (264)
Placebo	8.7 (89)	10.8 (89)	10.1 (79)	9.8 (257)
Total	10,5 (262)	13,5 (263)	11,5 (257)	

3.26 Interacțiunea între două variabile independente

Interacțiunea dintre două variabile independente se referă la diferențele în efectul unei variabile, în funcție de nivelul celei de-a doua variabile. De exemplu, poate un medicament produce efecte mai bune atunci când este combinat cu o dietă de scădere în greutate decât atunci când este combinat cu o dietă cu conținut limitat de sodiu. Este posibil să nu existe un efect semnificativ al acelui medicament atunci când toate grupurile de dietă sunt grupate, dar dacă ne uităm la efectele separat pentru fiecare grup de dietă, putem descoperi o interacțiune între cei doi factori: dietă și medicament.

Diagramele de mai jos ilustrează conceptul de efecte de interacțiune. WR înseamnă scădere în greutate, iar SR înseamnă restricție de sodiu (sare).



În exemplul 1, medicamentul B este mai bun decât medicamentul A la cei sub scădere în greutate, dar la cei cu restricție de sare, medicamentul A este mai bun decât medicamentul B. Dacă am compara doar media pentru medicamentul A, combinând diete, cu media pentru medicament B, ar trebui să spunem că nu există nicio diferență între medicamentul A și medicamentul B, dar dacă ne uităm la cele două diete separat, vedem efecte destul de diferite ale celor două medicamente.

În exemplul 2, nu există nicio diferență între cele două medicamente pentru cei care restricționează sare, dar există un efect mai mic al medicamentului A decât al medicamentului B pentru cei care slăbesc.

În exemplul 3, nu există nicio interacțiune; există un efect egal pentru ambele diete: cele două linii sunt paralele; pantele lor sunt aceleași. Medicamentul B este mai bun decât medicamentul A atât pentru cei care slăbesc, cât și pentru restricții de sare.

3.27 Exemplu de ANOVA bidirecțional

În continuare este un tabel de date din studiul TAIM care arată rezultatele unei analize bidirecționale a variației:

ANOVA bidirecțională din datele TAIM

Sursă	DF	Suma ANOVA a pătratelor	Pătrat mediu	Valoarea F	Probabilitate
Grup de droguri	2	1776,49	888,25	13.68	.0001
Grup de dieta	2	1165,93	582,96	8,98	.0001
Medicament × dietă	4	214,50	53,63	0,83	.509
Eroare	773	50.185,46	64,93		

Rețineți că pătratul mediu al erorii aici este 64,93 în loc de 66,9 când am făcut analiza unidirecțională. Acest lucru se datorează faptului că am explicat o parte din variația erorii ca fiind datorată efectelor dietei și efectelor de interacțiune (am „eliminat" aceste efecte din variația erorii). Astfel, 64,93 reprezintă varianța datorată erorii pure sau varianța inexplicabilă. Acum putem folosi rădăcina pătrată a acesteia, care este 8,06, ca estimare a abaterii standard comune. Explicăm rezultatele după cum urmează: Există un efect semnificativ al medicamentului (p = .0001) și un efect semnificativ al dietei (p = .0001), dar nicio interacțiune a medicamentului cu dietă (p = .509).

Am făcut deja cele trei comparații pe perechi, prin teste t pentru diferența dintre două medii între medicamente (adică, placebo vs. diuretic, placebo vs. beta-blocant și diuretic vs. beta-blocant). Putem face același lucru pentru cele trei diete. Valorile lor medii sunt afișate mai jos:

Grup de dieta	*n*	Scădere medie a tensiunii arteriale diastolice	Abaterea standard
Reducere în greutate	263	13.5	8.3
Restricție de sodiu	257	11.5	8.3
Dieta obisnuita	262	10.5	8.0

Estimarea comună a sd = 8,06

Dacă am face teste t, am constata că reducerea în greutate este mai bună decât dieta obișnuită (p = .0000), dar restricția de sodiu nu arată nicio îmbunătățire semnificativă față de dieta obișnuită (p = .16).

Reducerea în greutate în comparație cu restricția de sodiu este, de asemenea, semnificativ mai bună cu p = .005, care este mult sub p = .017 cerut de procedura Bonferroni. (T pentru această comparație în perechi este 2,83, care este peste valoarea critică de 2,39.)

3.28 Testul Kruskal – Wallis pentru a compara mai multe grupuri

Analiza variației este valabilă atunci când variabila de interes este continuă , provine dintr-o distribuție normală, adică curba familiară în formă de clopot, iar variațiile în cadrul fiecăruia dintre grupurile comparate sunt în esență egale. De multe ori, însă, trebuie să ne confruntăm cu situații în care dorim să comparăm mai multe grupuri pe o variabilă care nu îndeplinește toate condițiile de mai sus. Acesta ar putea fi un caz în care putem spune că o persoană este mai bună decât alta, dar nu putem spune exact cât de mult mai bună. Într-un astfel de caz, am clasifica oamenii și am compara grupurile utilizând testul Kruskal– Wallis pentru a determina dacă este probabil ca toate grupurile să provină dintr-o populație comună. Acest test este analog cu analiza unidirecțională a variației, dar în loc să folosească scorurile originale, folosește *clasamentele* scorurilor. Se numește *test neparametric* . Acest test este disponibil în multe programe de calculator, dar un exemplu apare în Anexa 3 .

3.29 Asociere și cauza: Coeficientul de corelație

O clasă comună de probleme în acumularea și evaluarea dovezilor științifice este evaluarea asocierii a două variabile. Există o asociere între sărăcie și dependența de droguri? Stresul emoțional este asociat cu bolile cardiovasculare?

Pentru a determina asocierea, trebuie mai întâi să cuantificăm ambele variabile. De exemplu, stresul emoțional poate fi cuantificat utilizând un instrument adecvat

testul psihologic al stresului sau prin definirea, evaluarea și evaluarea clară pe o scară a factorului de stres în situația de viață a unui individ, în timp ce hipertensiunea arterială (definită ca o citire a tensiunii arteriale) poate fi considerată ca aspectul particular al bolii cardiovasculare care trebuie studiat. . Când variabilele au fost cuantificate, trebuie calculată o măsură a asocierii pentru a determina puterea relației. Una dintre cele mai comune măsuri de asociere este coeficientul de corelație, r, care este un număr derivat din date și care poate varia între —1 și +1. (Majoritatea pachetelor de calculatoare statistice vor calcula coeficientul de corelație , dar dacă doriți să o faceți singur, metoda de calcul apare în Anexa 4.) Când $r = 0$, înseamnă că nu există nicio asociere între cele două variabile. Un exemplu în acest sens ar putea fi corelația dintre tensiunea arterială și numărul de fire de păr de pe cap. Când $r = +1$, o corelație pozitivă perfectă, înseamnă că există o relație directă între cele două variabile: un individ care are un scor mare la o variabilă are și un scor mare la cealaltă, iar scorul la o variabilă poate fi exact prezis din scorul celeilalte variabile. Acest tip de corelație există doar în modelele deterministe, unde există într-adevăr o relație funcțională. Un exemplu ar putea fi corelația dintre vârsta unui copac și numărul de inele pe care le are. Un coeficient de corelație de — 1 indică o relație inversă perfectă, unde un scor mare pentru o variabilă înseamnă un scor scăzut pentru cealaltă și unde, ca în corelația pozitivă perfectă, nu există nicio eroare de măsurare. Coeficienții de corelație între 0 și +1 și între 0 și — 1 indică diferite forțe ale asociilor.

Acești coeficienți de corelație se aplică atunci când relația de bază dintre cele două variabile este liniară. Luați în considerare un grup de persoane pentru fiecare dintre care avem o măsurare a greutății față de înălțime; vom constata că putem trage o linie dreaptă prin puncte. Există o asociere liniară între greutate și înălțime, iar coeficientul de corelație ar fi pozitiv, dar mai mic de 1. Când variabilele sunt continue, corelația calculată este corelația produs-moment Pearson. Dacă variabilele sunt clasate și ordonate în funcție de rang, calculăm corelația rang-ordine Spearman, care este o statistică neparametrică. Statisticile neparametrice sunt utilizate atunci când datele nu trebuie să fie distribuite normal și sunt ordinale (adică, pot fi sortate în ordine, dar distanțele dintre oricare două valori nu trebuie să fie aceleași). Un

Un exemplu este nivelul educațional, care poate fi clasificat în mai puțin de un liceu, a absolvit liceul, o anumită facultate, a absolvit o facultate și a primit o diplomă de absolvent. Acestea le puteți atribui numere de la 1 la 5, dar numerele nu reprezintă ani de studii, ci mai degrabă categorii de educație care sunt ordonate de la cea mai mică la cea mai mare categorie și le puteți clasifica diferit dacă doriți.

Diagramele din figura 3.9 ilustrează reprezentări ale diferiților coeficienți de relație.

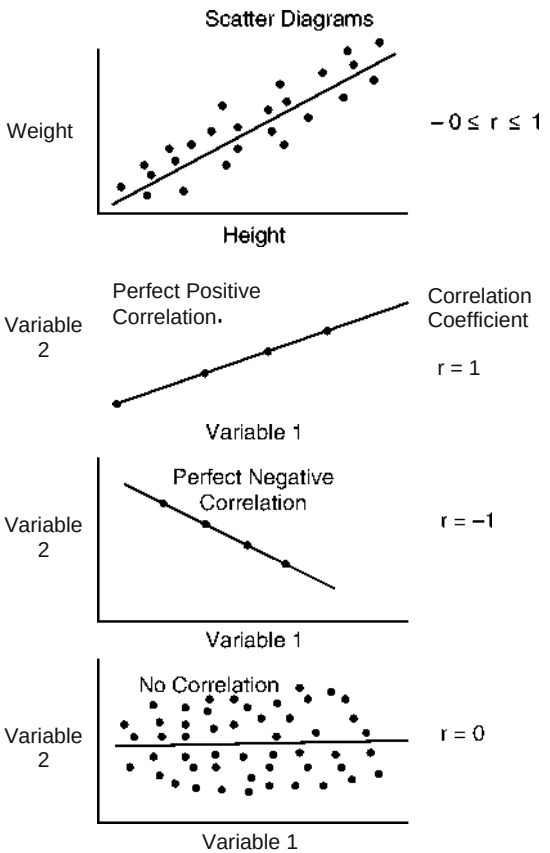


Figure 3.9 Correlations between 2 variables

3.30 Câteva puncte de reținut despre corelație

- Un coeficient de corelație pătrat (r^2) vă spune ce proporție a varianței unei variabile este explicată de cealaltă variabilă. Astfel, r = .40 înseamnă că .4 × .4 = .16 sau 16 % din variația unei variabile este explicată de cealaltă variabilă, dar restul de 84 % din variație se datorează altor factori.
- Cât de sus este mare? Răspunsul la această întrebare depinde de domeniul de aplicare, precum și de mulți alți factori, inclusiv precizia măsurării, precum și de faptul că pot exista relații diferite între două variabile în grupuri diferite de oameni. De exemplu, corelațiile dintre aptitudinea verbală și aptitudinea nonverbală, măsurate pentru școlari din Philadelphia prin teste naționale standardizate, variază de la .44 la .71, în funcție de rasă și clasa socială a grupurilor. [13]
- Mărimea coeficientului de corelație și semnificația statistică. Un coeficient de corelație foarte scăzut poate fi semnificativ statistic dacă dimensiunea eșantionului este suficient de mare. Ceea ce pare a fi un coeficient de corelație ridicat poate să nu atingă semnificația statistică dacă dimensiunea eșantionului este prea mică. Semnificația statistică vă spune că corelația pe care ați observat-o nu se datorează probabil întâmplării, dar nu vă spune nimic despre puterea asocierii.
- Corelația nu este cauzalitate. Acesta este cel mai important punct și unul care este adesea ignorat. Corelația nu vă spune nimic despre direcția relației și nici dacă există un al treilea factor care influențează ambele variabile. Cel mai cunoscut caz este cel care este invocat pentru a susține teoria conform căreia berzele aduc bebeluși. Există o corelație ridicată între numărul de accidente vasculare cerebrale dintr-o zonă și numărul de bebeluși născuți. Acest lucru a fost observat în orașul Copenhaga analizând înregistrările privind numărul de nașteri și numărul de berze pe o perioadă de 10 ani după al Doilea Război Mondial.[I] Se pare că după război, mai mulți oameni

[I]Referințe mulțumiri profesorului Angela Pignotti de la Modesta Junior College, Modesta California, Ornithologische Monatsberichte, 44 No. 2, Jahrgang, 1936, Berlin Ornithologische Monatsberichte, 48 No. 1, Jahrgang, 1940, Berlin Statistiches Jahrbuch Deutscher Gemeinden, Jahrgang3, 27–40. , 1932–1938, Gus tav Fischer, Jena.

s-a mutat la Copenhaga (deci s-au născut mai mulți bebeluși acolo) și mai multe construcții se desfășurau în oraș (deci berzele aveau mai multe locuri de cuibărit, deci mai multe berze). O parodie amuzantă pe această temă este de Thomas Hofer.[II]

3.31 Căi cauzale

Dacă obținem o corelație semnificativă, atunci ne întrebăm ce situații ar putea fi responsabile pentru aceasta. Figura 3.10 ilustrează câteva posibile relații structurale care pot sta la baza unui coeficient de corelație semnificativ. [14] Luăm în considerare două variabile, W (creșterea în greutate) și B (tensiunea arterială), iar r_{WB} reprezintă corelația dintre ele. Rețineți că numai în diagramele (1), (2) și (6) corelația dintre W și B apare din cauza unei relații cauzale între cele două variabile. În diagrama (1), W determină în întregime mine B; în diagrama (2), W este o cauză parțială a lui B; în diagrama (6), W este unul dintre câțiva determinanți ai lui B. În toate celelalte relații structurale, corelația dintre W și B apare din cauza influențelor comune asupra ambelor variabile. Astfel, trebuie subliniat că existența unei corelații între două variabile nu implică neapărat cauzalitate. Pot apărea corelații deoarece o variabilă este cauza parțială a alteia sau cele două variabile corelate au o cauză comună. Alți factori, cum ar fi eșantionarea , variația celor două populații și așa mai departe, afectează și dimensiunea coeficientului de corelație. Prin urmare, trebuie avută grijă în interpretarea acestor coeficienți.

[II]Noi dovezi pentru teoria berzei, Thomas Ho¨ fera, Hildegard Przyrembelb și Silvia Verlegerc, Institutul Federal pentru Evaluarea Riscurilor, Berlin, Biroul Comitetului Național pentru Alăptare la BfR, Berlin și Moașa Independentă, Berlin, Germania, Editura Blackwell Ltd. Epidemiologie pediatrică și perinatală 2004, 18, 88–92

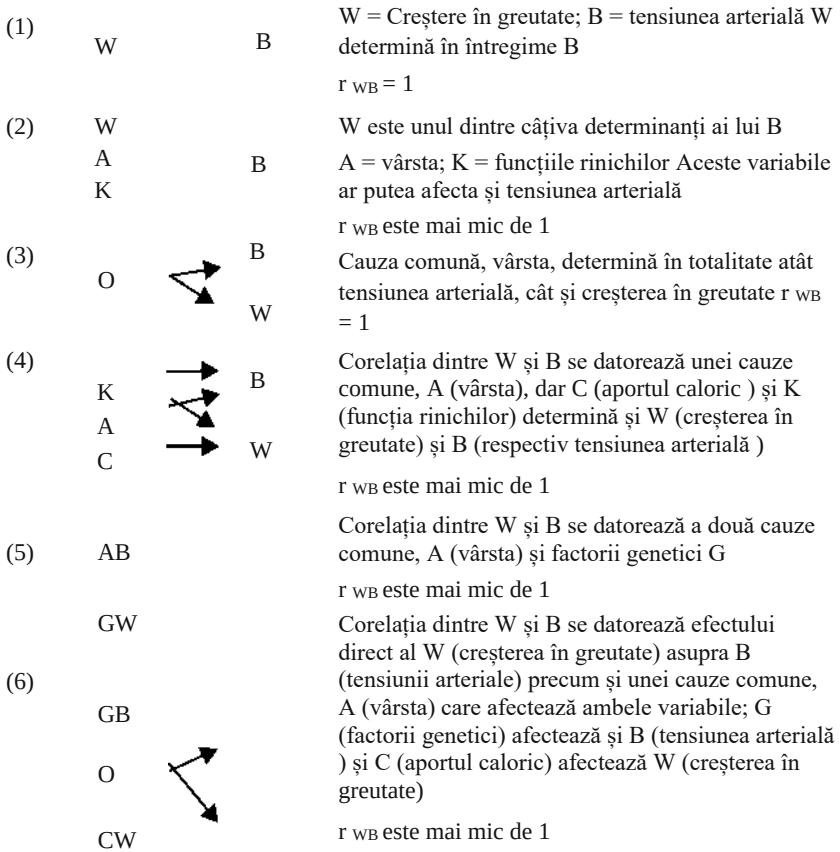

(1)
W
B
W = Creștere în greutate; B = tensiunea arterială W determină în întregime B
r WB = 1
(2)
W
A
K
B
W este unul dintre câțiva determinanți ai lui B
A = vârsta; K = funcțiile rinichilor Aceste variabile ar putea afecta și tensiunea arterială
r WB este mai mic de 1
(3)
O
B
W
Cauza comună, vârsta, determină în totalitate atât tensiunea arterială, cât și creșterea în greutate r WB = 1
(4)
K
A
C
B
W
Corelația dintre W și B se datorează unei cauze comune, A (vârsta), dar C (aportul caloric) și K (funcția rinichilor) determină și W (creșterea în greutate) și B (respectiv tensiunea arterială)
r WB este mai mic de 1
(5)
AB
GW
Corelația dintre W și B se datorează a două cauze comune, A (vârsta) și factorii genetici G
r WB este mai mic de 1
Corelația dintre W și B se datorează efectului direct al W (creșterea în greutate) asupra B (tensiunii arteriale) precum și unei cauze comune, A (vârsta) care afectează ambele variabile; G (factorii genetici) afectează și B (tensiunea arterială) și C (aportul caloric) afectează W (creșterea în greutate)
(6)
GB
O
CW
r WB este mai mic de 1

Figura 3.10 Relații posibile care stau la baza corelațiilor

3.32 Regresia

Rețineți că în figura 3.9 am trasat linii care par să se potrivească cel mai bine cu punctele de date. Acestea se numesc linii de regresie. Ele au următoarea formă Y = a + bX : În graficul de dispersie de sus etichetat (a), *Y* este variabila dependentă greutatea și X, sau înălțimea, este variabila independentă. spunem noi

acea greutate este o funcție de înălțime. Mărimea a este interceptarea. Este locul în care linia traversează axa Y. Mărimea b este panta și este rata de modificare a lui Y pentru o modificare unitară a lui X. Dacă panta este 0, înseamnă că avem o dreaptă paralelă cu axa x, ca în ilustrația (d). De asemenea, înseamnă că nu putem prezice Y din cunoașterea lui X, deoarece nu există nicio relație între Y și X. Dacă avem situația prezentată în grame dispersate (b) sau (c), știm exact cum se schimbă Y atunci când X se schimbă și putem prezice perfect Y dintr-o cunoaștere a lui X fără eroare. În diagrama de dispersie (a), putem vedea că pe măsură ce X crește, Y crește, dar nu putem prezice Y perfect deoarece punctele sunt împrăștiate în jurul dreptei pe care am trasat-o. Putem, totuși, să punem limite de încredere în jurul predicției noastre , dar mai întâi trebuie să stabilim forma liniei pe care ar trebui să o trasăm prin puncte. Trebuie să estimăm valorile pentru interceptarea și panta. Acest lucru se face prin găsirea „liniei cele mai potrivite".

Linia care se potrivește cel mai bine punctelor are următoarele caracteristici: dacă luăm fiecare dintre punctele de date și îi calculăm distanța verticală față de linie și apoi pătratăm acea distanță, suma acelor distanțe pătrate va fi mai mică decât suma acestor distanțe pătrate de la orice altă linie pe care am putea-o trasa. Aceasta se numește potrivirea celor mai mici pătrate. Luați în considerare datele de mai jos, unde Y ar putea fi un scor la un test și X ar putea fi un scor la un alt test.

	Scor	
Individual	*X*	*Y*
O	5	7
B	8	4
C	15	8
D	20	10
E	25	14

Calculele pentru determinarea liniei celei mai potrivite sunt prezentate în Anexa 4 . Cu toate acestea, majoritatea pachetelor de calculatoare statistice pentru computerele personale oferă un program de regresie liniară care face aceste calcule. Figura 3.11 ilustrează aceste puncte reprezentate într-un scattergram și arată linia celor mai mici pătrate.

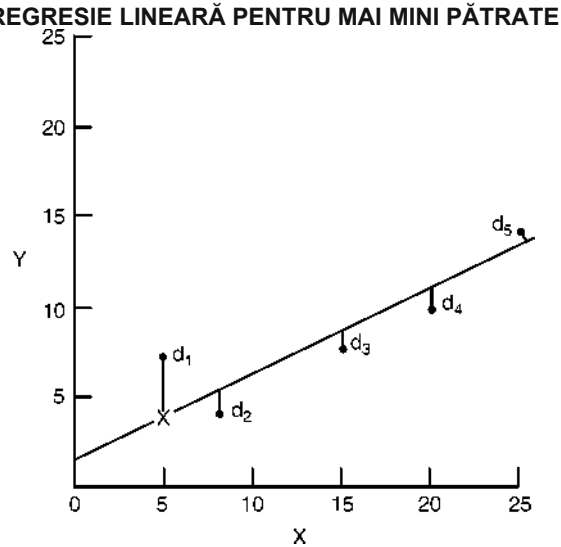


Figura 3.11 Linia de regresie a celor mai mici pătrate

Ecuația dreptei este $Y = 2{,}76 + .40\,X$. Intercepta *a este 2,76* , astfel încât linia traversează y axa la $Y = 2{,}76$. Panta este de .40. De exemplu, putem calcula un Y prezis pentru $X = 10$ pentru a obține

$$Y = 2 : 76 + (\ .\ 40)(10) = 2 : 76 + 4 = 6 : 76$$

D i_s sunt distanțe de la puncte la linie. Suma acestor distanțe pătrate este mai mică pentru această linie decât ar fi pentru orice altă linie pe care am putea-o trasa.

Coeficientul de corelație pentru aceste date este .89. Pătratul coeficientului de corelație, r^2, poate fi interpretat ca proporția varianței în Y *care* este explicată prin X. În exemplul nostru, $.89^2 = .79$; astfel, 79 % din variația lui Y este explicabilă prin variabila X , iar 21 % nu este contabilizată.

3.33 Legătura dintre regresia liniară și coeficientul de corelație

Coeficientul de corelație și panta dreptei de regresie liniară sunt legate prin

formule

$$r = b\frac{s_x}{s_y}, \qquad b = r\hat{}\frac{s_x}{s_x}$$

unde s_x este abaterea standard a variabilei X, s_y este abaterea standard a variabilei Y, b este panta dreptei și r este coeficientul de corelație.

3.34 Regresia liniară multiplă

Când avem două sau mai multe variabile independente și o variabilă dependentă continuă, putem folosi analiza regresiei multiple. Forma pe care o ia aceasta este

$$Y = a + b_i X_i + b_2 X_2 + b_3 X_3 + \ldots + b_x X_k$$

De exemplu, Y poate fi tensiunea arterială și X_1 poate fi vârsta, X_2 poate fi greutatea și X_3 poate fi antecedente familiale de hipertensiune arterială. Putem avea cât mai multe variabile, unde ultima variabilă este a k-a variabilă. B_is sunt coeficienți de regresie. Rețineți că antecedentele familiale de hipertensiune arterială nu sunt o variabilă continuă. Poate fi fie da, fie nu. Numim aceasta variabilă dihotomică și o putem folosi ca orice altă variabilă într-o ecuație de regresie, atribuind un număr fiecăruia dintre cele două răspunsuri posibile , de obicei făcând un răspuns da = 1 și un răspuns nu = 0.

Urmează un exemplu din studiul TAIM și este menit doar să vă ofere o idee despre cum să interpretați o ecuație de regresie multiplă. Această analiză se referă la grupul de 89 de persoane care au fost alocate unui medicament placebo și unui regim de scădere în greutate. Variabila dependentă este modificarea tensiunii arteriale.

Variabilele independente sunt prezentate mai jos:

Variabilă	Coeficient: b_i	p
Intercepta	- 15.49	.0016
Vârstă	.077	.359
Cursa 1 = negru 0 = nenegru	4.22	.021
Sex 1 = masculin 0 = feminin	1,50	.390
Lire sterline pierdute	.13	.003

Notă: Sexul este codificat ca 1 dacă este bărbat și 0 dacă este femeie; rasa este codificată ca 1 dacă este neagră și 0 dacă nu este neagră; p este folosit pentru a testa dacă coeficientul este semnificativ diferit de 0. Ecuația, atunci, este

modificarea tensiunii arteriale =

$$\wedge 15{:}49 \wedge + \wedge{:}077 \wedge (\text{age}) + 4{:}22_{(\text{race})} + {}_{15(\text{sex})} + {:}^{13}{}_{(\text{Change in weight})}$$

Vârsta nu este semnificativă ($p = .359$) și nici sexul ($p = .390$). Cu toate acestea, rasa este semnificativă ($p = 0{,}021$), indicând faptul că negrii au avut o probabilitate mai mare decât cei non- negri să aibă o scădere a tensiunii arteriale, controlând simultan toate celelalte variabile din ecuație. Kilogramele pierdute sunt, de asemenea, semnificative, ceea ce indică faptul că cu cât pierderea în greutate era mai mare, cu atât scăderea tensiunii arteriale era mai mare.

Modelele de regresie liniară presupun că observațiile sunt independente unele de altele, adică sunt necorelate. În exemplul de mai sus, modificarea greutății pentru o altă persoană este independentă de modificarea greutății pentru o altă persoană din eșantion. Dar în studiile longitudinale, unde indivizii au măsuri repetate, observațiile sunt corelate deoarece sunt observații asupra acelorași persoane. Aici este adecvat să se utilizeze o tehnică de regresii pe date corelate numită GEE, care reprezintă ecuații de estimare generalizate, descrisă într-un articol de Hanley și colegii săi. [15]

3.35 Rezumat Până acum

Investigarea unei probleme științifice necesită adesea o analiză statistică, mai ales

acolo unde există variabilitate în ceea ce privește caracteristicile de interes. Variabilitatea poate apărea din două surse: caracteristica poate fi variabilă în mod inerent în populație și/sau poate exista o eroare de măsurare.

În acest capitol, am subliniat că pentru a evalua un program sau un medicament, pentru a compara două grupuri pe o anumită caracteristică și pentru a efectua o investigație științifică a oricărei probleme, este necesară cuantificarea variabilelor.

Variabilele pot fi cuantificate ca fiind discrete sau continue și există tehnici statistice adecvate care se ocupă de fiecare dintre acestea. Am luat în considerare aici testul chi-pătrat, intervalele de încredere, testul Z, testul t, analiza variației, corelația și regresia. Am subliniat că în testarea ipotezelor, suntem supuși la două tipuri de erori: eroarea de a respinge o ipoteză atunci când de fapt este adevărată și eroarea de a accepta o ipoteză atunci când de fapt este falsă. Scopul unui studiu bine conceput este de a minimiza probabilitatea de a face aceste tipuri de erori. Statisticile nu vor înlocui un design experimental bun, dar este un instrument necesar pentru a evalua dovezile științifice obținute din studii bine concepute.

Din punct de vedere filozofic, statistica este o reflectare a vieții în două - privințe importante: (1) ca și în viață, nu putem fi niciodată siguri de nimic (dar în statistică putem pune o cifră de probabilitate care descrie gradul de incertitudine) și (2) totul este un compromis — în statistică, între certitudine și precizie sau între două tipuri de erori; în viață, ei bine, completează-ți propriile compromisuri.

Capitolul 4
MAI MULT DESPRE EPIDEMIOLOGIE

Medicina pentru a produce sănătate trebuie să examineze boala; iar muzica pentru a crea armonie trebuie să investigheze discordia.

Plutarh 46–120
d.Hr

4.1 Utilizările epidemiologiei

Epidemiologia poate fi definită ca studiul distribuției sănătății și bolilor în grupuri de oameni și studiul factorilor care influențează această distribuție. Epidemiologia modernă cuprinde, de asemenea, evaluarea modalităților diagnostice și terapeutice și furnizarea de servicii de îngrijire a sănătății . Există o progresie în procesul științific (de-a lungul dimensiunii creșterii credibilității dovezilor), de la observația întâmplătoare, la formarea de ipoteze, la observația controlată, la studii experimentale. Figura 4.1 este o reprezentare schematică a utilizărilor epidemiologiei. Instrumentele utilizate în acest demers sunt în provincia epidemiologiei și biostatisticii . Tehnicile utilizate în aceste discipline permit „detectivilor medicali" să descopere o problemă medicală, să evalueze dovezile despre cauzalitatea sau etiologia acesteia și să evalueze intervențiile terapeutice pentru combaterea problemei.

Epidemiologia descriptivă oferă informații despre tiparul bolilor, despre „cine are ce și cât din el", informații care sunt esențiale pentru planificarea asistenței medicale și alocarea rațională a resurselor. Astfel de informații pot descoperi adesea modele de apariție care sugerează relații etiologice și pot duce la strategii preventive. Astfel de studii sunt de obicei de tip transversal și conduc la formarea de ipoteze care pot fi apoi testate în studii caz-control, prospective și experimentale. Studiile clinice și alte tipuri de studii controlate servesc la evaluarea modalităților terapeutice și a altor mijloace de intervenție și astfel determină în cele din urmă standardele de practică medicală, care la rândul lor au impact asupra deciziilor de planificare a îngrijirii sănătății. În secțiunea următoare, vom lua în considerare concepte epidemiologice selectate.

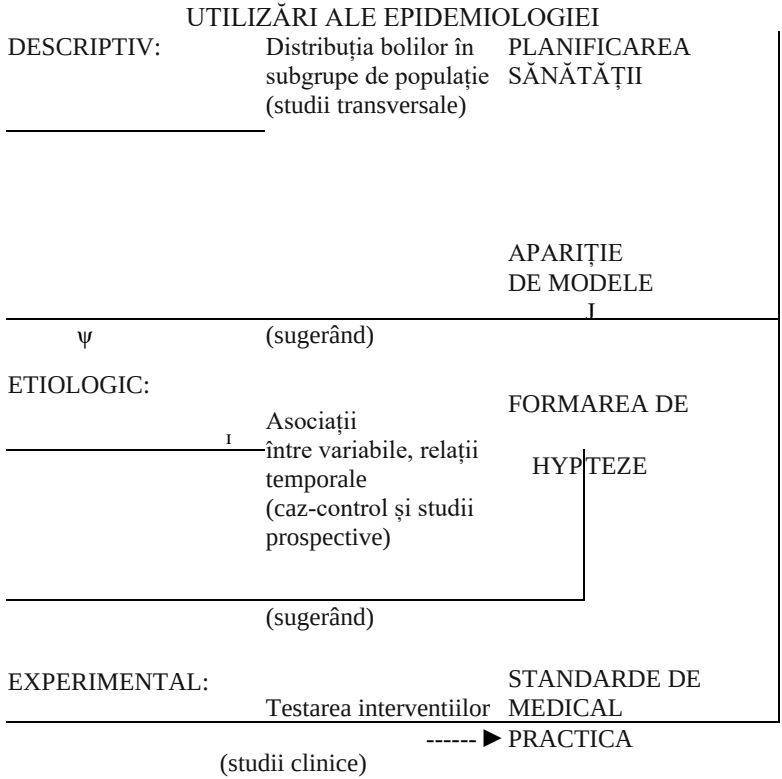


Figura 4.1 Utilizări ale epidemiologiei

4.2 Câteva concepte epidemiologice: rate de mortalitate

În 1900, cele trei cauze majore de deces erau gripa sau pneumonia, tuberculoza și gastroenterita. Astăzi, cele trei cauze majore de deces sunt bolile de inimă, cancerul și accidentele; al patrulea este accidente vasculare cerebrale. Decesele cauzate de accident vascular cerebral au scăzut dramatic în ultimele decenii, probabil datorită controlului îmbunătățit al hipertensiunii arteriale, unul dintre factorii de risc primari pentru accident vascular cerebral. Aceste modele în schimbare de mortalitate reflectă condițiile de mediu în schimbare, o trecere de la bolile acute la cele cronice și o

îmbătrânirea populației supusă bolilor degenerative. Știm acest lucru dintr-o analiză a ratelor .

Compararea ratelor definite între diferite subgrupuri de indivizi poate oferi indicii cu privire la existența unei probleme de sănătate și poate conduce la specificarea condițiilor în care această problemă de sănătate identificată este probabil să apară și să înflorească.

La utilizarea tarifelor, trebuie reținute următoarele puncte:

(1) O rată este o proporție care implică un numărător și un numitor.
(2) Atât numărătorul, cât și numitorul trebuie să fie clar definite, astfel încât să știți la ce grup (numitor) se referă rata dvs.
(3) Numătorul este conținut în (este o submulțime a) numitorului. Acest lucru este în contrast cu un raport în care numărătorul se referă la un grup diferit de numitor.

Ratele de mortalitate se referă la numărul de decese care au loc într-un anumit subgrup de populație și oferă adesea unul dintre primele indicii ale unei probleme de sănătate. Următoarele definiții sunt necesare înainte de a continua discuția:

Rata brută anuală a mortalității (sau rata mortalității) este:

Numărul total de decese
pe parcursul unui an în populația expusă riscului
Populația expusă riscului
(de obicei luată ca populație la mijlocul anului)

Cauza - Rata de mortalitate anuală specifică este:

Numărul de decese survenite din cauza unei anumite cauze
în cursul anului în populația expusă riscului
Populația expusă riscului (toți cei în viață la jumătatea anului)

Rata de mortalitate anuală specifică de vârstă este:

$$\frac{\text{Numărul de decese survenite la grupa de vârstă dată în cursul anului în populația cu risc}}{\text{Populația expusă riscului (toți cei în viață la mijlocul anului)}}$$

Un motiv pentru a lua populația la mijlocul anului ca numitor este că o populație poate crește sau scădea în timpul anului în cauză, iar populația de la mijlocul anului este o estimare a numărului mediu pe parcursul anului. Se poate vorbi totuși de rate de mortalitate pe o perioadă de 5 ani, mai degrabă decât pe o perioadă de 1 an, și se poate defini populația expusă riscului ca fiind toți cei în viață la începutul perioadei.

4.3 Tarife ajustate în funcție de vârstă

Atunci când se compară ratele de mortalitate între două populații, trebuie să se țină seama de componența pe vârstă a populațiilor. Deoarece persoanele în vârstă au un număr mai mare de decese la 1.000 de persoane, dacă o populație este puternic ponderată de persoanele în vârstă, rata brută a mortalității ar fi mai mare decât la o populație mai tânără, iar o comparație între cele două grupuri ar putea reflecta doar discrepanța de vârstă, mai degrabă. decât o diferență intrinsecă în experiența mortalității. O modalitate de a face față acestei probleme este de a compara ratele de deces specifice vârstei, ratele de deces specifice unui anumit grup de vârstă. O altă modalitate care este utilă atunci când este necesară o cifră rezumată generală este utilizarea ratelor ajustate în funcție de vârstă . Acestea sunt rate ajustate la ceea ce ar fi dacă cele două populații comparate ar avea aceleași distribuții de vârstă ca o populație standard selectată în mod arbitrar.

De exemplu, tabelul de mai jos arată ratele de mortalitate brute și ajustate în funcție de vârstă pentru Statele Unite în cinci perioade de timp 15.7. Ajustarea se face la distribuția pe vârstă a populației în 1940 precum și la distribuția pe vârstă a populației în 2000. Astfel, vedem că în 1991, rata ajustată în funcție de vârstă a fost de 5,1/1.000 la ajustarea la standardul din 1940, dar rata brută a mortalității a fost de 8,6/1.000. Aceasta înseamnă că dacă în 1991 distribuția pe vârstă a populației ar fi aceeași ca în 1940, atunci rata mortalității ar fi fost de doar 5,1/1.000 de persoane. Brutul și

Ratele ajustate în funcție de vârstă pentru 1940 sunt aceleași, deoarece populația din 1940 servește ca populație „standard" a cărei distribuție pe vârstă este utilizată ca bază pentru ajustare.

Când a fost ajustată la standardul anului 2000, rata ajustată în funcție de vârstă a fost de 9,3. Dacă în 1991 distribuția pe vârstă ar fi aceeași ca în 2000, atunci rata mortalității ar fi fost de 9,3/1.000 de persoane. Deci, ratele ajustate în funcție de vârstă depind de populația standard utilizată pentru ajustare. Rețineți că rata ajustată în funcție de vârstă pe baza populației în anul 2000 este mai mare decât rata ajustată în funcție de vârstă pe baza populației din 1940; acest lucru se datorează faptului că populația este mai în vârstă în anul 2000.

An	Rata brută a mortalității la 1.000 de persoane	Rata ajustată în funcție de vârstă (la populație în 1940)	Rată ajustată în funcție de vârstă (la populație în 2000)
1940	10.8	10.8	17.9
1960	9.5	7.6	13.4
1980	8.8	5.9	10.4
1991	8.6	5.1	9.3
2001	8.5	Nu se calculează după 1998	8.6

Deși atât ratele brute, cât și cele ajustate în funcție de vârstă au scăzut din 1940, scăderea ratei ajustate în funcție de vârstă este mult mai mare. Modificarea procentuală a mortalității brute între 1940 și 1991 a fost (10,8–8,6)/10,8 = 20,4 %, în timp ce modificarea procentuală a ratei ajustate în funcție de vârstă a fost (10,8-5,1)/10,8 = 0,528 sau 52,8 %.

Motivul pentru aceasta este că populația este în creștere. De exemplu, proporția persoanelor de 65 de ani și peste s-a dublat între 1920 și 1960, crescând de la 4,8 % din populație în 1920 la 9,6 % în 1969. După 1998, Centrul Național de Statistică a Sănătății a folosit populația din 2000 ca populație standard. față de care s-au făcut ajustări. Rata brută și rata mortalității ajustată în funcție de vârstă în anul 2001 sunt similare, și asta pentru că distribuția pe vârstă din 2001 este similară cu cea din 2000, astfel încât ajustarea vârstei nu schimbă cu adevărat rata mortalității.

Ratele ajustate în funcție de vârstă sunt numere fictive – nu vă spun câți oameni au murit efectiv la 1.000, ci câți ar fi murit .

dacă compoziţiile de vârstă ar fi aceleaşi la cele două populaţii. Cu toate acestea, ele sunt adecvate în scopuri de comparație. Metodele de efectuare a ajustării vârstei sunt descrise în Anexa 5 .

4.4 Incidență și prevalență

Prevalența și incidența sunt două măsuri de morbiditate (boală).

Prevalența unei boli este definită ca:

Numărul de persoane cu o boală
Numărul total de persoane din populație
expuse riscului la un anumit moment în timp

(Acest lucru este cunoscut și sub denumirea de prevalență punctuală , dar mai general denumită doar „prevalență.") De exemplu, prevalența hipertensiunii arteriale în 1973 în rândul bărbaților de culoare, cu vârsta cuprinsă între 30 și 69 de ani, definită ca o tensiune arterială diastolică (DBP) de 95 mmHg sau mai mult la un program de depistare a tensiunii arteriale realizat de Programul de detectare și urmărire a hipertensiunii arteriale (HDFP), [16] a fost calculat a fi

4.268 *cu DBP* > 95 *mmHg*
15.190 *de bărbați de culoare cu vârsta* de 30 de ani — 69 *examinați*
× 100 = 28 . 1 la 100

Mai multe puncte trebuie remarcate cu privire la această definiție:

(1) Grupul de risc (numitorul) este clar definit ca fiind bărbați de culoare, cu vârste cuprinse între 30 și 69 de ani.
(2) Momentul de timp este specificat ca timp de screening.
(3) Definiția bolii este clar specificată ca o tensiune arterială diastolică de 95 mmHg sau mai mare. (Acesta poate include persoanele care sunt tratate pentru boală, dar a căror presiune este încă mare și cei care nu sunt tratați.)
(4) Numărătorul este subgrupul de indivizi din grupul de referință (numitorul) care îndeplinesc definiția bolii.

Prevalența poate fi, de asemenea, ajustată în funcție de vârstă la o populație

standard, ceea ce înseamnă că estimările prevalenței sunt ajustate la ceea ce ar fi dacă populația de interes ar avea aceeași distribuție pe vârstă ca o populație standard (a se vedea Anexa 5 și Secțiunea 4.3). Prevalența hipertensiunii estimată de Hispanic Community Health Study/Study of Latinos (HCHS/SOL), [17] în 2007-2011, a fost găsită a fi de 25,5% atunci când vârsta ajustată la standardul anului 2000. Hipertensiunea arterială a fost definită ca tensiune arterială sistolică de 140 mmHg sau mai mult sau tensiune arterială diastolică de 90 mmHg sau mai mult sau pe medicamente pentru hipertensiune arterială.

Incidența este definită ca:

$$\frac{\text{Numărul de cazuri noi ale unei boli pe unitatea de timp}}{\text{Numărul total la risc la începutul acestei perioade de timp}}$$

De exemplu, studiile au descoperit că incidența de 10 ani a unui eveniment coronarian major (cum ar fi atacul de cord) în rândul bărbaților albi, cu vârste cuprinse între 30 și 59 de ani, cu tensiunea arterială diastolică de 105 mmHg sau mai mare, la momentul în care au fost observați pentru prima dată, a fost găsită. să fie 183 la 1.000. [18] Aceasta înseamnă că dintre 1.000 de bărbați albi, cu vârsta cuprinsă între 30 și 59 de ani, care aveau tensiune arterială diastolică peste 105 mmHg la începutul perioadei de observație de 10 ani, 183 dintre ei au avut un eveniment coronarian major (atac de cord sau moarte subită) în timpul următorii 10 ani. Printre bărbații albi cu tensiune arterială diastolică < 75 mmHg, incidența pe 10 ani a unui eveniment coronarian a fost de 55/1.000. Astfel, compararea acestor două rate de incidență, 183/1.000 pentru cei cu tensiune arterială mare față de 55/1.000 pentru cei cu tensiune arterială scăzută, poate duce la deducerea că tensiunea arterială crescută este un factor de risc pentru boala coronariană.

Adesea, se poate auzi cuvântul „incidență” folosit atunci când ceea ce înseamnă cu adevărat este prevalență. Ar trebui să aveți grijă de o astfel de utilizare incorectă. De exemplu, s-ar putea să auzi sau chiar să citești într-un jurnal medical că incidența diabetului zaharat în 1973 a fost de 42,6 la 1.000 de persoane, cu vârste cuprinse între 45 și 64 de ani, când ceea ce înseamnă cu adevărat este că prevalența a fost de 42,6/1.000. Lucrul de reținut este că prevalența se referă la existența unei boli într-o anumită perioadă de timp, în timp ce incidența se referă la cazuri noi ale unei boli care se dezvoltă într-o anumită perioadă de timp.

Rețineți că rata mortalității este incidență, în timp ce morbiditatea poate fi exprimată ca incidență sau prevalență. Într-o boală cronică, prevalența este mai mare decât incidența, deoarece prevalența include atât cazuri noi, cât și cazuri existente care ar fi putut să fi apărut pentru prima dată cu mult timp în urmă, dar pacienții afectați au continuat să trăiască cu această afecțiune. Pentru o boală care este fie rapid fatală, fie rapid vindecată, incidența și prevalența pot fi similare. Prevalența poate fi stabilită prin efectuarea unui sondaj sau a unui screening al unei populații țintă și numărarea cazurilor de boală existente la momentul anchetei. Acesta este un studiu transversal. Cifrele de incidență sunt mai greu de obținut decât cifrele de prevalență, deoarece pentru a stabili incidența, trebuie să se identifice un grup de persoane care nu suferă de boala în cauză (adică o cohortă), să le observe pe o perioadă de timp și să se determine câte dezvoltă boala de-a lungul timpului. acea perioadă de timp. Implementarea unui astfel de proces este dificilă și costisitoare.

4.5 Rata de mortalitate standardizată

Rata standardizată de mortalitate (SMR) este raportul dintre numărul de decese observate și numărul de decese așteptate. Numărul așteptat pentru o anumită grupă de vârstă, de exemplu, este adesea obținut din statisticile populației.

$$\text{SMR} \quad \frac{\text{decese observate}}{\text{decese așteptate}}$$

4.6 Persoană-Ani de observație

Ocazional, se vede o rată prezentată ca un anumit număr de evenimente pe persoană-ani de observație, mai degrabă decât pe numărul de indivizi observați într-o anumită perioadă de timp. Pe persoană-ani (sau luni) este utilă ca unitate de măsură atunci când oamenii sunt observați pentru diferite perioade de timp. Să presupunem că observați cohorte de oameni fără boli de inimă pentru a determina dacă incidența bolilor de inimă este mai mare pentru fumători decât pentru cei care renunță. Cei care părăsesc trebuie definiți, de exemplu,

precum cei care au renunțat cu mai mult de 5 ani înainte de începerea observației.

S-ar putea defini diferit persoanele care renunță și ar putea obține rezultate diferite, așa că este important să specificați definiția. Alte considerații includ controlul pentru durata de timp fumată, care ar fi o funcție de vârstă la începutul fumatului și vârsta la începutul perioadei de observare, numărul de țigări fumate și așa mai departe. Dar, pentru simplitate, vom presupune că toți dintre fumători au fumat o cantitate egală și toți cei care au renunțat au fumat o cantitate egală înainte de a renunța.

Putem exprima rata de incidență a bolilor de inimă într-o anumită unitate de timp, să zicem 10 ani, ca numărul care au dezvoltat boala în acel timp, împărțit la numărul de persoane pe care le-am observat (număr cu risc). Totuși, să presupunem că nu i-am observat pe toți pentru aceeași perioadă de timp. Acest lucru s-ar putea întâmpla deoarece unele persoane s-au mutat sau au murit din alte cauze sau au fost înscrise în studiu în momente diferite sau din alte motive. Într-un astfel de caz, am putea folosi ca numitor numărul de ani-persoană de observație.

De exemplu, dacă individul 1 a fost înscris la momentul 0 și a fost observat timp de 4 ani, apoi a pierdut din urmărire, el/ea ar fi contribuit cu 4 ani-persoană de observație. Zece astfel de indivizi ar contribui cu 40 de ani-persoană de observație. Un alt individ observat timp de 8 ani ar fi contribuit cu 8 persoane-ani de observație, iar 10 astfel de indivizi ar fi contribuit cu 80 persoană-ani de observație pentru un total de 120 persoană-ani. Dacă s-au dezvoltat șase cazuri de boli de inimă printre cei observați, rata ar fi de 6 la 120 de persoane-ani, mai degrabă decât de 6/10 indivizi observați. Rețineți că, dacă numitorul este dat ca ani persoană, nu știți dacă se referă la 120 de persoane observate fiecare timp de un an, sau la 12 persoane observate fiecare timp de 10 ani sau o combinație. O altă problemă cu această metodă de exprimare a ratelor este că reflectă experiența medie de-a lungul intervalului de timp, dar este posibil ca rata bolilor de inimă să fie aceeași pentru fumători ca și pentru cei care renunță la fum în primii 3 ani, iar ratele încep să se separe după că. În orice caz, sunt disponibile diverse metode statistice pentru a fi utilizate cu analiza persoană-an.

4.7 Variabile dependente și independente

În studiile de cercetare, dorim să cuantificăm relația dintre un set de variabile, pe care le putem considera predictori sau determinanți, și o variabilă de rezultat sau criteriu de care suntem interesați. Această variabilă de rezultat, pe care obiectivul nostru este să o explicăm, este variabila dependentă.

O variabilă dependentă este un factor a cărui valoare depinde de nivelul altui

factor, care este denumit variabilă independentă . În exemplul fumatului de țigară și al mortalității prin cancer pulmonar, durata și numărul de țigări fumate sunt variabile independente de care depinde mortalitatea prin cancer pulmonar (astfel, mortalitatea prin cancer pulmonar este variabila dependentă).

4.8 Tipuri de studii

In Sectiunea 1.4 , am descris diferite tipuri de modele de studiu, în contextul discuției noastre despre metoda științifică și despre modul în care știm ceea ce știm. Acestea au fost studii observaționale, care pot fi transversale, caz-control sau studii prospective și experimentale, care sunt studii clinice. În secțiunile următoare, vom lua în considerare tipurile de inferențe care pot fi derivate din datele obținute din aceste modele diferite.

Obiectivul este de a evalua relația dintre un factor de interes (variabila independentă), pe care uneori îl vom numi expunere și o variabilă de rezultat (variabila dependentă).

Studiile observaționale se disting prin momentul în care se fac măsurători pe variabilele dependente și independente, așa cum este ilustrat mai jos. În studiile transversale, atât variabilele dependente, cât și cele independente (rezultat și expunere) sunt măsurate în același timp, în prezent. În studiile caz-control, rezultatul este măsurat acum și expunerea este estimată din trecut. În studiile prospective, expunerea (variabila independentă) este măsurată acum, iar rezultatul este măsurat în viitor. În secțiunea următoare, vom discuta diferitele inferențe care trebuie făcute din studii transversale versus studii prospective.

	Timpul de măsurare		
	Trecut	Prezent	Viitor
Secțiune transversală		Rezultatul expunerii	
Caz-control	Rezultatul expunerii		
Prospective		Rezultatul expunerii	

4.9 Analize transversale versus longitudinale asupra datelor

Studiile prospective sunt uneori cunoscute și ca studii longitudinale, deoarece oamenii sunt urmăriți longitudinal, în timp. Examinarea datelor longitudinale poate duce la inferențe destul de diferite decât cele care se obțin din analizele transversale ale datelor. De exemplu, luați în considerare vârsta și tensiunea arterială.

Studiile transversale au arătat în mod repetat că tensiunea arterială sistemică medie este mai mare în fiecare grup de vârstă succesiv de 10 ani, în timp ce presiunea diastolică crește pentru grupele de vârstă până la 50 de ani și apoi atinge un platou. Din aceste tipuri de studii nu se poate spune că tensiunea arterială crește odată cu vârsta pentru că presiunile măsurate la bărbații de 30 de ani, de exemplu, nu au fost obținute asupra acelorași indivizi 10 ani mai târziu când aveau 40 de ani, ci au fost obținute pentru un set diferit de bătrâni de 40 de ani. Pentru a determina efectul vârstei asupra tensiunii arteriale, ar trebui să aruncăm o privire longitudinală sau prospectivă asupra acelorași indivizi pe măsură ce îmbătrânesc. O interpretare a curbei observate pentru tensiunea arterială diastolică, de exemplu, ar putea fi aceea că indivizii de peste 60 de ani care aveau presiuni diastolice foarte mari au murit, lăsând în viață doar acei indivizi cu presiune mai scăzută suficient de mult pentru a fi incluși în eșantionul celor care aveau sânge. presiunea măsurată în aspectul în secțiune transversală.

Diagramele din Figura 4.2 ilustrează impactul posibil al unui „efect de cohortă”, o vedere în secțiune transversală și o vedere longitudinală a acelorași date. (Literele indică grupuri de persoane examinate într-un anumit an.)

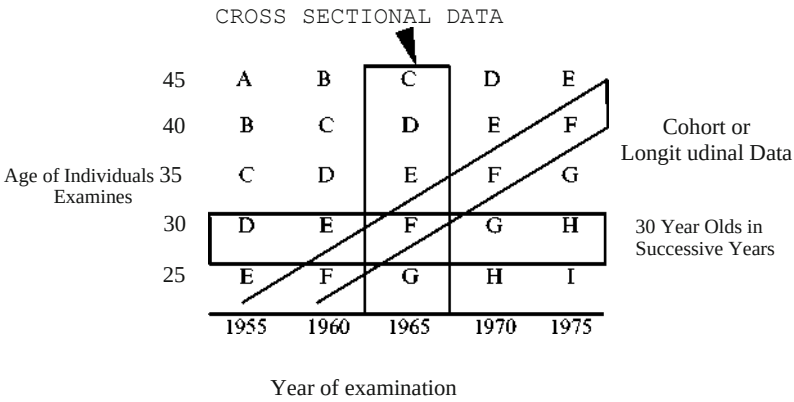

CROSS SECTIONAL DATA
45 A B C D E
40 B C D E F
35 C D E F G
30 D E F G H
25 E F G H I
1955 1960 1965 1970 1975
Age of Individuals Examines
Cohort or Longit udinal Data
30 Year Olds in Successive Years
Year of examination

Figure 4.2

Dacă luați tensiunea arterială a tuturor grupelor în 1965 și comparați grupul F cu grupul D, veți avea o comparație transversală a persoanelor de 30 de ani cu persoanele de 40 de ani la un moment dat. Dacă comparați grupul F în 1965 cu grupul F (aceiași indivizi) în 1975, veți avea o comparație longitudinală. Dacă comparați grupul F în 1965 cu grupul H în 1975, veți avea o comparație a tensiunii arteriale la persoanele de 30 de ani în diferite momente în timp (o privire orizontală).

Aceste comparații pot duce la concluzii destul de diferite, așa cum se arată în exemplele schematice din Figura 4.3 folosind numere fictive pentru a reprezenta tensiunea arterială diastolică medie.

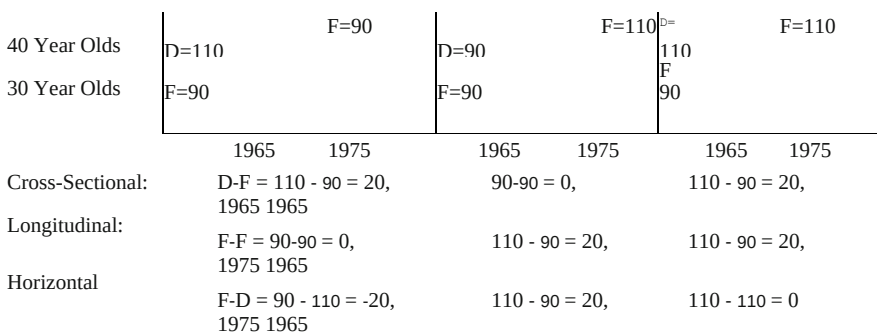


Figura 4.3 Comparații transversale versus longitudinale

În exemplul (1), măsurătorile din 1965 indică faptul că tensiunea arterială diastolică medie pentru persoanele de 30 de ani (grupa F) a fost de 90 mmHg, iar pentru persoanele de 40 de ani (grupa D) a fost de 110 mmHg. Privind grupul F 10 ani mai târziu, când aveau 40 de ani, indică tensiunea arterială diastolică medie a fost de 90 mmHg. Rezultă următoarele calcule:

Aspect în secțiune transversală	D - F = 110 - 90 = 20 1965 1965
Concluzie	Persoanele de 40 de ani au tensiune arterială mai mare decât persoanele de 30 de ani (cu 20 mmHg)
Aspect longitudinal	F - F ¼ 90 - 90 ¼ 0 1975 1965
Concluzie	Tensiunea arterială nu crește odată cu vârsta
Aspect orizontal (comparații de cohorte)	F - F ¼ 90 - 110 = - 20 1975 1965
Concluzie	Oamenii de 40 de ani în 1975 au tensiune arterială mai mică decât cei de 40 de ani în 1965

O posibilă interpretare: tensiunea arterială nu crește odată cu vârsta, dar forțe de mediu diferite au funcționat pentru cohorta F decât pentru cohorta D.

În exemplul (2), avem

Aspect în secțiune transversală	D - F = 90 - 90 = 0 mmHg 1965 1965
Concluzie	Din datele transversale, concluzionăm că tensiunea arterială nu este mai mare odată cu vârsta înaintată
Aspect longitudinal	F - F = 110 - 90 = 20 1975 1965

Concluzie	Din datele longitudinale, concluzionăm că tensiunea arterială crește odată cu vârsta
Aspect orizontal	F - D = 110 - 90 = 20 1975 196
Concluzie	Oamenii de 40 de ani în 1975 au o tensiune arterială mai mare decât cei de 40 de ani în 1965

O posibilă interpretare : tensiunea arterială crește odată cu vârsta și diferiți factori de mediu operați pe cohorta F decât pe cohorta D.

În exemplul (3), avem

Aspect în secțiune transversală	D - F ¼ 110 - 90 ¼ 20 1965 1965
Concluzie	În secțiune transversală, a existat o creștere a tensiunii arteriale pentru persoanele de 40 de ani față de cea pentru cei de 30 de ani.
Aspect longitudinal	F - F ¼ 110 - 90 ¼ 20 1975 1965
Concluzie	Longitudinal, se vede că tensiunea arterială crește odată cu vârsta
Aspect orizontal	F - D = 110 - 110 = 0 1975 1965
Concluzie	Nu a existat nicio modificare a tensiunii arteriale în rândul persoanelor de 40 de ani pe parcursul perioadei de 10 ani

O posibilă interpretare: tensiunea arterială crește odată cu vârsta (susținută atât de date longitudinale, cât și de date transversale), iar factorii de mediu afectează ambele cohorte în mod similar.

4.10 Măsuri ale riscului relativ: concluzii din studii prospective (studiul Framingham)

În studiile epidemiologice, suntem adesea interesați să știm cât de mult mai probabil este un individ să dezvolte o boală dacă este expus la un anumit factor decât individul care nu este atât de expus. O măsură simplă a unei astfel de probabilități se numește risc relativ (RR). Este raportul dintre două estimări de incidență: rata de dezvoltare a bolii pentru persoanele cu factor de expunere, împărțită la rata de dezvoltare a bolii pentru persoanele fără factor de expunere . Să presupunem că dorim să determinăm efectul hipertensiunii arteriale (hipertensiunea) asupra dezvoltării bolilor cardiovasculare (BCV). Pentru a obține riscul relativ, trebuie să calculăm ratele de incidență. Putem folosi datele dintr-un studiu prospectiv clasic, Framingham Heart Study. [19]

Acesta a fost un studiu epidemiologic prospectiv de pionierat al unui eșantion de populație din orășelul Framingham, Massachusetts. Începând cu 1948, o cohortă de oameni a fost selectată pentru a fi urmărită bienal. Termenul de cohortă se referă la un grup de indivizi urmăriți longitudinal pe o perioadă de timp. O cohortă de naștere, de exemplu, ar fi populația de indivizi născuți într-un anumit an. Cohorta Framingham a fost un eșantion de persoane alese la începutul perioadei de studiu și a inclus bărbați și femei cu vârsta cuprinsă între 30 și 62 de ani la începutul studiului. Acești indivizi au fost observați pe o perioadă de 20 de ani și au fost determinate morbiditatea și mortalitatea asociate bolilor cardiovasculare . S-a obținut o fișă standardizată a spitalului și un certificat de deces, examinarea clinică a fost repetată la intervale de 2 ani, iar preocuparea majoră a studiului Framingham a fost evaluarea relației dintre caracteristicile determinate la persoanele bine și dezvoltarea ulterioară a bolii.

Prin acest studiu au fost identificați „factori de risc" pentru bolile cardiovasculare. Factorii de risc sunt caracteristicile fiziologice antecedente sau obiceiurile alimentare și de viață, a căror prezență crește probabilitatea - individului de a dezvolta boli cardiovasculare la un moment dat. Printre cei mai importanți factori predictivi identificați în studiul Framingham au fost creșterea tensiunii arteriale, creșterea colesterolului seric și fumatul. Glicemia crescută și rezultatele anormale ale electrocardiogramei de repaus sunt, de asemenea, predictive pentru viitoarele boli cardiovasculare.

Riscul relativ poate fi determinat prin următorul calcul:

Incidența bolilor cardiovasculare (cazuri noi)
pe o anumită perioadă de timp în rândul persoanelor fără
BCV la începutul perioadei de studiu care
au factorul de risc în cauză (de exemplu , hipertensiune arterială)
Incidența bolilor cardiovasculare într-o anumită perioadă de timp în rândul
persoanelor lipsite de boli cardiovasculare inițial , care nu prezintă riscul
factorul în cauză (tensiune arterială normală)

Din datele Framingham, calculăm pentru bărbați în studiu

RR of CVD within 18 years after first exam = 353 . 2 / 10 000 de persoane cu risc cu hipertensiune arterială certă 123 . 9 / 10 000 de persoane cu risc, fără hipertensiune arterială

$$\frac{353.2}{123.9} = 2.85$$

Aceasta înseamnă că un bărbat cu hipertensiune arterială certă are de 2,85 ori mai multe șanse de a dezvolta BCV într-o perioadă de 18 ani decât un bărbat care nu are hipertensiune arterială. Pentru femei, riscul relativ este

$$\frac{187.9}{57.3} = 3.28$$

Aceasta înseamnă că hipertensiunea arterială prezintă un risc relativ oarecum mai mare pentru femei. Dar rețineți că riscul absolut pentru persoanele cu hipertensiune arterială certă (adică, incidența BCV) este mai mare la bărbați decât la femei, fiind de 353,2 la 10.000 de bărbați față de 187,9 la 10.000 de femei.

Estimările de incidență prezentate mai sus au fost ajustate în funcție de vârstă. Ajustarea vârstei este discutată în Secțiunea 4.3 . Adesea, cineva poate dori să se adapteze pentru alte variabile, cum ar fi starea de fumat, diabetul, nivelul de colesterol și alți factori care pot fi, de asemenea, legați de rezultat. Acest lucru poate fi

realizat prin analiza de regresie logistică multiplă și prin analiza pericolelor proporționale Cox, care sunt descrise în secțiuni. 4.17 și , respectiv, 4.19 , dar mai întâi vom descrie cum riscul relativ poate fi calculat din studii prospective sau estimat din studii caz-control.

4.11 Calculul riscului relativ din studii prospective

Riscul relativ poate fi determinat direct din studii prospective prin construirea unui tabel 2 × 2 după cum urmează [20] :

		BOALA (în curs de dezvoltare în perioada specificată)		
		Yes	No	
RISK FACTOR (determined at beginning of study period)	PRESENT (high blood pressure)	*a* = 90	*b* = 403	*a* + *b* = 493 (persons with factor)
	ABSENT (normal blood pressure)	*c* = 70	*d* = 1201	*c* + *d* = 1271 (persons without factor)

Riscul relativ este

$$\frac{\text{incidența bolii printre cei cu TA mare}}{\text{incidența bolii printre cei cu TA normală}} = \frac{a / (a + b) \_ 90 / 493}{c / (c + d) \„ 70 / 1271} = 3.31$$

Riscul relativ sau raportul hazardului poate fi calculat din modelele de regresie a hazardelor proporționale Cox (care permit ajustarea pentru alte variabile), așa cum este descris în Secțiunea 4.19 .

4.12 Odds Ratio: Estimarea riscului relativ din studii de caz – control

Un studiu caz-control este unul în care investigatorul încearcă să stabilească o asociere între prezența unei caracteristici (un factor de risc) și apariția unei boli, pornind de la un eșantion de persoane bolnave și un grup de control de persoane nebolnave și prin observarea prezenței sau absenței caracteristicii în fiecare dintre aceste două grupuri. Poate fi ilustrat prin construirea unui tabel 2 × 2 după cum urmează:

		BOALA	
		PREZENT	ABSENT
RISC FACTOR	PREZENT	*o*	*b*
	ABSENT	*c*	*d*
		a + *c*	*b* + *d*

(numar de persoane (numar de persoane *cu* boala) *fara* boala)

Obiectivul este de a determina dacă proporția persoanelor cu boală care au factorul este mai mare decât proporția persoanelor fără boală care au factorul. Cu alte cuvinte, se dorește să știm dacă a /(a + c) este mai mare decât b /(b + d).

Studiile caz-control sunt adesea denumite studii retrospective deoarece investigatorul trebuie să adune date despre variabilele independente retrospectiv. Variabila dependentă – prezența bolii – este obținută în momentul prelevării, spre deosebire de studiile prospective în care variabilele independente sunt măsurate în momentul prelevării, iar variabila dependentă este măsurată într-un moment viitor (adică, când boala se dezvoltă). Adevărata distincție între studiile caz-control sau retrospective și studiile prospective are de-a face cu selectarea indivizilor pentru studiu - cei cu și fără boală în studiile caz-control/retrospective și cei cu și fără factorul de interes în studiile prospective.

Deoarece în studiile prospective eșantionăm persoanele cu caracteristica de interes și persoanele fără caracteristica , putem obține

riscul relativ direct prin calcularea ratelor de incidență a bolii în aceste două grupuri. În studiile caz-control, totuși, eșantionăm pacienți cu și fără boală și

notăm prezența sau absența caracteristicii de interes în aceste două grupuri; nu avem o măsură directă a incidenței bolii. Cu toate acestea, făcând anumite ipoteze , putem face unele aproximări ale riscului relativ dacă am putea măsura ratele de incidență printr-un studiu prospectiv. Aceste aproximări sunt valabile cel mai bine pentru bolile cu incidență scăzută.

Pentru a estima riscul relativ din studiile caz-control, rețineți că

$$a = \{ a + Ü) a \{ c + d)$$
$$c = \{ c + d) c \{ a + \text{b})$$

Acum presupuneți că, deoarece boala nu este atât de comună, *c* este neglijabilă în raport cu *d* (cu alte cuvinte, printre persoanele fără factor de risc, nu sunt atât de mulți care vor face boala, raportat la numărul de persoane cine nu o va primi). Să presupunem, de asemenea, că, *în populație* , *a* este neglijabilă în raport cu *b* , deoarece chiar și printre persoanele cu factor de risc, nu atât de mulți vor face boala în comparație cu numărul celor care nu o vor face. Apoi, termenii din paranteze devin *d* la numărător și *b* la numitor, astfel încât

$$a\{c + d) \quad , \quad .ad$$
$$—\{— + \wedge \text{ se reduce la } OR = —$$

Aceasta este cunoscută sub numele de *odds ratio* (OR) și este o bună estimare a riscului relativ atunci când boala este rară.

Un exemplu despre modul în care este calculat cota de cote este prezentat mai jos. Într-un studiu caz-control al cancerului pulmonar, a fost obținut tabelul de mai jos. [21] Rețineți că aici nu eșantionăm fumători și nefumători. Mai degrabă, eșantionăm pe cei cu și fără boală. Deci, deși în *populația generală a* este mic în raport cu *b* , în acest eșantion nu este așa.

	Pacienți cu cancer pulmonar		Controale corelate cu alte boli	
Fumători de 15-24 de țigări pe zi	475	*o*	431	*b*
nefumători	7	*c*	61	*d*
	(persoane cu boala)		(persoane fără boala)	

Raportul de șanse, ca estimare a riscului relativ de a dezvolta cancer pulmonar pentru persoanele care fumează 15-24 de țigări pe zi, în comparație cu nefumătorii este

$$\textit{Odds ratio} = \frac{475 \times 61}{431 \times 7} = 9.60 = \textit{estimate cf relativerisk}$$

Aceasta înseamnă că fumătorii de 15-24 de țigări pe zi au de 9,6 ori mai multe șanse de a face cancer pulmonar decât nefumătorii.

Încă un lucru despre raportul de șanse este că este raportul dintre șansele de cancer pulmonar pentru cei care fumează 15-24 de țigări pe zi, comparativ cu șansele de cancer pulmonar pentru cei care nu fumează . În exemplul de mai sus, obținem

475
pentru fumători: șansele de cancer pulmonar sunt

7
pentru nefumători: șansele de cancer pulmonar sunt

. 475 / 431
cote ratiocf = — 7 / 61 —

Deci, ideea este că raportul de șanse este raportul de șanse, indiferent dacă boala este rară sau nu. Este întotdeauna raportul dintre șansele de îmbolnăvire pentru cei cu

expunerea față de șansele de îmbolnăvire pentru cei fără expunere. Dar atunci

când boala este rară, este și o bună estimare a riscului relativ.

Putem, de asemenea, să punem limite de încredere pentru raportul de cote. Am calculat cota ca fiind 9,60. Limitele de încredere de 95 % pentru un raport de cote (OR) sunt

$$SAU \; x \; e \; [\; i1 : 96 \; p \; ai\text{^}]$$

Limita superioară de încredere de 95 % =

$$SAU \; X \; e \; [\; 1 \; 96 \; p \; 475 \; \text{^} \; 43i \; \text{^} \; 1 \; ⅛]$$

$$SAU \; X \; e \; (1 \; 96 \; x : 405) = 9 \; . \; 6 \; X \; e \; 794 = 9 : 6 \; x \; 2 : 21 = 21 : 2$$

Limită inferioară de încredere de 95 % =

$$SAU \; x \; e \; (-1 \; 9 \; 6 \; x : 405) = 9 : 6 \; x \; e \; - \; 7 \; 94 = 9 \; . \; 6 \; x \; . \; 45 = 4 \; . \; 3$$

Notă: $e \; . \; 79 \; 4 = -I- = .45$

$e : 794$

Astfel, *intervalul de încredere* este 4,3–21,2.

Adesea exprimăm acest lucru ca (OR; limite de încredere de 95 %), care este exemplul nostru este:

$$(\; OR = 9{,}60 \; ; \; 95 \; \% \; CI : 4{,}3 \; , \; 21{,}2 \;)$$

Ratele de cote pot fi calculate din regresia logistică (care permite ajustarea pentru alte variabile) așa cum este descris în Secțiunea 4.17 .

4.13 Risc atribuibil

Riscul atribuibil (RA) este:

Riscul la persoanele expuse — risc la persoanele neexpuse :

Riscul atribuibil populației (PAR) este:

Prevalența factorului de risc AR X

În timp ce riscul relativ se referă la riscul unei boli la persoanele expuse în raport cu riscul la persoanele neexpuse, riscul atribuibil se referă la diferența de risc absolut al persoanelor expuse comparativ cu persoanele neexpuse. Ne poate spune cât de mult exces de risc există din cauza expunerii la cei expuși. În exemplul din Secțiunea 4.11 , riscul pe 10 ani în rândul celor cu hipertensiune arterială a fost 90/493 = 0,183 (sau 183 la 1.000 de persoane cu hipertensiune arterială), în timp ce la cei cu tensiune arterială normală, acesta a fost de 70/1.271. = .055 (sau 55 la 1.000 cu presiune normală).

Astfel, riscul atribuibil la cei expuși (adică, cu hipertensiune arterială) este .183-.055 = .128 (128 la 1.000). Cu alte cuvinte, evenimentele de boli de inima la 128 din cele 183 de persoane la 1.000 cu hipertensiune arterială pot fi atribuite hipertensiunii arteriale. De asemenea, putem exprima acest exces ca procent din riscul expusului care este atribuit expunerii

$$\frac{128/1000}{183/1000} \qquad 128 = .70 \textit{ or } 70\%$$

Dar, trebuie să fim foarte atenți cu privire la o astfel de atribuire - este valabilă numai atunci când putem presupune că expunerea cauzează boala (după ce luăm în considerare confuzie și alte surse de părtinire).

Riscul atribuibil populației (PAR) este o măsură utilă atunci când dorim să vedem cum am putea reduce morbiditatea sau mortalitatea prin eliminarea unui factor de risc. Depinde de prevalența factorului de risc în populație, așa cum sa menționat mai sus. Iată un exemplu din Inițiativa pentru sănătatea femeilor (descris mai detaliat în Capitolul 6). Într-un studiu clinic s-a descoperit că femeile aflate în postmenopauză care luau estrogen plus progestativ au avut o rată anuală a bolii coronariene de 39 la 10.000, comparativ cu o rată de 33 la 10.000 pentru femeile care luau placebo. [22]

Astfel,

$$AR \qquad \frac{39 - 33}{10,\ 000} = .0006$$

sau șase evenimente de boală coronariană în exces la 10.000 de femei care iau acest preparat, pe an.

Deoarece aproximativ 6 000 000 de femei luau acel preparat hormonal la momentul respectiv (expuse), atunci 0,0007 × 6 000 000 = 3 600 de evenimente

coronare de boală cardiacă pe an ar putea fi atribuite luării de estrogen plus progestativ.

Prevalența utilizării estrogenului plus progestativ estimată din același studiu când a fost început a fost de aproximativ 18 %. Dacă folosim această estimare,

PAR = AR × prevalența factorului de risc = : 0006 × : 18 = : 000108;

Astfel, dacă s-ar elimina utilizarea de estrogen plus progestativ, ar fi cu 10,8 la 100.000 de femei în postmenopauză mai puține care au avut evenimente de boli de inimă.

4.14 Bias de răspuns

Există multe tipuri diferite de părtinire care v-ar putea conduce fie să subestimați, fie să supraestimați dimensiunea unui risc relativ de cote și este important să încercați să anticipați potențialele surse de părtinire și să le evitați. Ilustrația de pe pagina următoare arată impactul unui tip de posibilă prejudecată: prejudecata de constatare sau de răspuns.

Să presupunem că aveți următoarea situație. Din 100 de persoane expuse la un factor de risc, 20% dezvoltă boala și din 100 de persoane neexpuse, 16% dezvoltă boala, rezultând un risc relativ de 1,25, așa cum se arată în ilustrație.

Acum imaginați-vă că doar 60% dintre cei expuși răspund la urmărire sau sunt constatați că au sau nu boala, o rată de răspuns de 60% în rândul celor expuși . Mai presupunem că toți cei care nu răspund se întâmplă să fie printre cei care nu dezvoltă boală . Riscul relativ ar fi calculat ca 2,06.

Acum imaginați-vă că doar 60% din răspunsul neexpus, o rată de răspuns de 60% în rândul celor neexpuși și toți cei neexpuși care nu răspund se întâmplă să fie printre cei care nu au boala. Acum, estimarea riscului relativ este de .75.

Pentru a rezuma, puteți obține estimări contradictorii ale riscului relativ dacă aveți rate de răspuns diferențiate. Prin urmare, este foarte important să obțineți un răspuns sau o constatare cât mai completă posibil. Urmează tabelele care prezintă aceste calcule.

FULL RESPONSE RATE

EXPOSURE	D I S E A S E +	–	
	R = 100% 20	100% 80	100
	100% 16	100% 84	100
	36	164	200

$$RR = \frac{20/100}{16/100} = \frac{.20}{.16} = 1.25$$

DIFFERENTIAL RESPONSE RATE

EXPOSURE	DISEASE +	DISEASE –	
+	R = 100% 20	50% 40	60 (response rate = 60%)
–	100% 16	100% 84	100
	36	124	160

RR = *20160* = .33 = 2.06
16 / 100 .16

DIFFERENTIAL RESPONSE RATE

EXPOSURE	DISEASE +	DISEASE –	
+	R = 100% 20	100% 40	100
–	100% 16	52% 84	60 (response rate = 60%)
	36	124	160

RR = 20 / 100. = .20 = .75

16 / 60 .266

4.15 Variabile de confuzie

O variabilă de confuzie este una care este strâns asociată atât cu variabila independentă, cât și cu rezultatul de interes pentru cei neexpuși. De exemplu, o variabilă de confuzie în studiile de cafea și boli de inimă poate fi fumatul. Deoarece unii băutori de cafea sunt și fumători, dacă un studiu a descoperit o relație între consumul de cafea (variabila independentă) și dezvoltarea bolilor de inimă (variabila dependentă), ar putea însemna cu adevărat că fumatul este cel care provoacă boli de inimă, mai degrabă decât cafea. În acest exemplu, fumatul este variabila de confuzie.

Dacă atât variabila de confuzie, cât și variabila independentă de interes sunt strâns asociate cu variabila dependentă, atunci relația observată între variabilele independente și dependente poate fi într-adevăr o reflectare a relației adevărate dintre variabila de confuzie și variabila dependentă. O modalitate intuitivă de a privi acest lucru este să ne imaginăm că, dacă o variabilă de confuzie ar fi perfect asociată cu o variabilă independentă, ar putea fi înlocuită cu ea. Este important să se țină seama sau să se ajusteze pentru variabilele confuze în proiectarea și analiza statistică a studiilor pentru a evita inferențe greșite.

Există mai multe abordări pentru a trata potențialii factori de confuzie. O abordare este de a trata aceasta în proiectarea studiului prin potrivire, de exemplu, așa cum este descris în Secțiunea 4.16 de mai jos; un alt mod de a controla variabilele confuze este în faza de analiză a datelor, prin utilizarea analizei multivariate, așa cum este descris în secțiunile de mai jos. O discuție excelentă se găsește în Modern Epidemiology de Rothman, Lash și Greenland.

4.16 Potrivire

O abordare pentru a trata potențialii factori de confuzie este de a potrivi subiecții din cele două grupuri pe variabila de confuzie. În exemplul discutat mai sus cu privire la studiile despre cafea și bolile de inimă, am putea potrivi subiecții în istoricul lor de fumat, deoarece fumatul poate fi un factor de confuzie al relației dintre cafea și bolile de inimă. Ori de câte ori am înscris un băutor de cafea în studiu, am stabili dacă persoana respectivă era fumătoare. Dacă pacientul ar fi fumător, următorul pacient care ar fi

înscriși, care nu a fost un băutor de cafea (adică, un membru al grupului de comparație) ar trebui să fie și un fumător. Pentru fiecare nefumător care consumă cafea, va fi înscris un nefumător care nu consumă cafea. În acest fel, am avea același număr de fumători în cele două loturi. Aceasta este cunoscută sub numele de potrivire unu-la-unu. Există și alte modalități de a se potrivi, iar acestea sunt discutate mai pe larg în cărțile lui Anderson, Elwood și Rothman, așa cum este menționat în secțiunea Lecturi sugerate.

În studiile caz-control, găsirea unui grup de comparație adecvat poate fi dificilă. De exemplu, să presupunem că un investigator studiază efectul cafelei asupra cancerului pancreatic. Anchetatorul alege ca martori pacienti in spital in acelasi timp si in aceeasi sectie cu cazurile dar cu alt diagnostic decat cancer. Este posibil ca pacienții internați pentru alte probleme gastro-intestinale decât cancerul să fi renunțat voluntar la consumul de cafea pentru că le-a deranjat stomacul. Într-o astfel de situație, obiceiurile de consum de cafea ale celor două grupuri ar putea fi similare, iar investigatorul ar putea să nu găsească o asociere mai mare a consumului de cafea cu cazurile decât cu martorii. Un grup mai adecvat ar putea fi pacienții dintr-o secție diferită, să zicem o secție de ortopedie. Dar aici ar trebui să fie atenți să se potrivească în funcție de vârstă, deoarece pacienții ortopedici pot fi mai tineri decât în celelalte cazuri dacă spitalul se întâmplă să fie într-o zonă de schi, de exemplu, unde schiul imprudent duce la picioare rupte, sau pot fi în mod substanțial. mai în vârstă decât celelalte cazuri dacă sunt mulți pacienți cu proteze de șold din cauza căderilor la vârstnici sau a osteoartritei.

Trebuie subliniat că factorul care este potrivit nu poate fi evaluat în ceea ce privește relația cu rezultatul. Astfel, dacă comparăm două grupuri de femei pentru efectul aportului de vitamina A asupra cancerului de col uterin și facem un studiu caz-control în care înregistrăm cazuri de cancer de col uterin și controale potrivite în funcție de vârstă, nu vom putea spune din acest studiu dacă vârsta este legată de cancerul de col uterin. Acest lucru se datorează faptului că ne-am asigurat că distribuțiile de vârstă sunt aceleași atât în grupul de caz, cât și în grupul de control, prin potrivirea în funcție de vârstă, deci, evident, nu vom putea găsi diferențe de vârstă între grupuri.

Unii statisticieni consideră că potrivirea se face adesea inutil și că, dacă aveți un studiu suficient de mare, randomizarea simplă sau randomizarea stratificată este adecvată pentru a asigura o distribuție echilibrată a factorilor de confuzie. În plus, metodele de analiză multivariată, cum ar fi
regresia logistică sau modelele de pericole proporționale oferă o altă modalitate, de obicei mai bună, de a controla factorii de confuzie. O bună discuție despre

potrivire poate fi găsită în cartea Methods in Observational Epidemiology , de Kelsey și colegii.

4.17 Regresie logistică multiplă

Analiza de regresie logistică multiplă este utilizată pentru a calcula probabilitatea ca un eveniment să se întâmple în funcție de mai multe variabile independente. Este util în controlul factorilor de confuzie atunci când se examinează relația dintre o variabilă independentă și apariția rezultatului (de exemplu, cum ar fi atacul de cord) într-o anumită perioadă de timp. Ecuația ia forma

$$P(\text{eveniment}) = \frac{1}{1 + e^{-k}}$$

$$\text{unde } k = C_{\text{ø}} + C_1 X_1 + C_2 X_2 + C_3 X_3 + \ldots + C_m X_m$$

$$(\text{rețineți că } e^{-k} = 1/e^{k})$$

Fiecare X_i este o anumită variabilă independentă și coeficienții corespunzători, C ' s, sunt calculați din datele obținute în studiu. De exemplu, să luăm datele Framingham pentru probabilitatea ca un bărbat să dezvolte boli cardiovasculare în decurs de 8 ani. Boala cardiovasculară (BCV) a fost definită ca boală coronariană, infarct cerebral, claudicație intermitentă sau insuficiență cardiacă congestivă.

$$P(\text{CVD}) = \frac{1}{1 \; ^{\wedge} e^{[-19.77 + .37(\text{vârstă}) - .002(\text{vârstă})_2 + .026(\text{chl}) + .016(\text{SBP}) + .558(\text{SM}) + 1.053(\text{LVH}) + .602(\text{GI}) - .00036(\text{chl x vârsta})]}}$$

Unde

chl = colesterol seric.
SBP = tensiunea arterială sistolica .

SM = 1 dacă da pentru fumat, 0 dacă nu.
LVH = hipertrofie ventriculară stângă , 1 dacă da, 0 dacă nu.
Gl = intoleranță la glucoză , 1 dacă da, 0 dacă nu.

De exemplu, să presupunem că luăm în considerare un bărbat de 50 de ani al cărui colesterol este de 200 și tensiunea arterială sistolică este de 160, care fumează și care nu are LVH și nu are intoleranță la glucoză. Când înmulțim coeficienții cu valorile acestui individ asupra variabilelor independente și facem calculele necesare, obținem o probabilitate de .17. Aceasta înseamnă că acest individ are 17 șanse la 100 de a dezvolta o formă de boală cardio-vasculară pe o perioadă de 8 ani.

Coeficienții dintr-o analiză de regresie logistică multiplă pot fi utilizați pentru a calcula raportul de șanse pentru un factor, controlând în același timp toți ceilalți factori. Modul de a face acest lucru este de a lua logaritmul natural e ridicat la coeficientul pentru variabila de interes, dacă variabila este una dihotomică (adică, codificată ca 1 sau 0). De exemplu, șansele de apariție a bolilor cardiovasculare pentru fumători în comparație cu nefumătorii în rândul bărbaților, controlând în același timp vârsta, colesterolul, tensiunea arterială sistolica, hipertrofia ventriculară stângă și intoleranța la glucoză este de $55^{8} = 1{,}75$. Aceasta înseamnă că o persoană care fumează are un risc de 1,75 ori mai mare de a face BCV (în decurs de 8 ani) decât cel care nu fumează dacă acești doi indivizi sunt egali în raport cu celelalte variabile din ecuație. Acest lucru este echivalent cu a spune că riscul fumătorului este cu 75% mai mare decât al nefumătorilor.

Dacă vrem să comparăm șansele unei persoane cu o tensiune arterială sistolica de 200 față de cineva cu tensiunea arterială sistolica de 120, toți ceilalți factori fiind egali, o calculăm după cum urmează:

$$SAU = e^{\beta(\equiv\text{-}\iota 2o)} = e^{\equiv(8°)} = e^{\iota^{\wedge}} = 3,6$$

Bărbatul cu tensiune arterială sistolică de 200 mmHg are de 3,6 ori mai multe șanse de a dezvolta boală decât cel cu presiunea de 120. (Bineînțeles, acest lucru ar implica că cineva cu tensiunea arterială sistolică de 260 ar avea, de asemenea, de 3,6 ori mai multe șanse de a dezvolta BCV. decât unul cu presiune de 180. Dacă ipoteza unei creșteri liniare a riscului nu s -ar menține, atunci predicția ar fi incorectă.)

Regresia logistică poate fi utilizată în studiile caz-control. Creșterea lui e la coeficientul variabilei de interes ne oferă odds ratio. Intervalele de încredere pentru raportul de cote sunt

$$SAU = e^{[\beta - 1{,}96(s\,e)]} \text{ ca limită inferioară}$$

$$\text{SAU} = e^{[\beta + 1.96(s_e)]} \text{ ca limita superioara :}$$

se este eroarea standard a coeficientului beta și este o ieșire a tuturor pachetelor de calculatoare statistice.

Regresia logistică multiplă este adecvată pentru studiile transversale și caz-control atunci când variabila dependentă (rezultatul) este dihotomică (adică, poate fi codificat ca 1 = eveniment, 0 = niciun eveniment) și când întrebarea se referă la apariția evenimentului de interes într-o anumită perioadă de timp și toate persoanele sunt urmărite pentru acea perioadă de timp. Cu toate acestea, atunci când timpul de urmărire pentru persoanele din studiu diferă, atunci trebuie utilizată analiza de supraviețuire, așa cum este descris în Secțiunile 4.18 și 4.19 .

4.18 Analiza supraviețuirii: Metode de tabel de viață

Analiza de supraviețuire a datelor ar trebui să fie utilizată atunci când perioadele de urmărire diferă mult pentru diferite persoane sau când aceștia intră în studiu în momente diferite. Poate deveni destul de complex, iar această secțiune este destinată doar introducerii conceptelor. Să presupunem că doriți să comparați supraviețuirea pacienților tratați prin două metode diferite și să presupunem că aveți datele prezentate mai jos. [23] O vom analiza folosind curbele de supraviețuire Kaplan–Meier.

DECES LA O LUNĂ DATE ÎN DOUĂ GRUPE

Stare: (D = decedat în luna respectivă; L = trăind în luna respectivă)

(+ înseamnă că pacientul a fost pierdut la urmărire și a fost văzut ultima dată în viață în luna respectivă)

Stare:	DLDDD	DDDDL
Grupa A:	4, 5+ 9, 11, 12 Grupa B:	2, 3, 4, 5, 6+

În fiecare grup, patru pacienți au murit până la 12 luni, iar unul a fost văzut în viață ceva timp în acel an, așa că nu știm dacă acel pacient era mort sau în viață la sfârșitul anului. Dacă am privi datele în acest fel, ar trebui să spunem că supraviețuirea până la 1 an a fost aceeași în ambele grupuri.

	Grup	
La sfarsitul a 12 luni	**O**	**B**
Mort	4	4
În viață	1	1
Rata de supraviețuire	20%	20%

Cu toate acestea, o modalitate mai adecvată de a analiza astfel de date este prin curbele de supraviețuire. Punctele pentru curbe sunt calculate așa cum se arată în tabelul de mai jos.

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7
Caz#	Ora în Mos.	Stare	# puncte. intra	Prop. Mort mort intrat	Prop. surv. $P_1 = 1 - q_j$	Cum. surv. $P_1 = P_i - 1 \times Pi$
Grupa A						
1	4	Mort	5	1/5 = .2	.80	1 × .8 = .8
2	5	Surv	4	0/4 = .0	1.00	.8 × 1 = .8
3	9	Mort	3	1/3 = .33	.67	.8 × .67 = .53
4	11	Mort	2	1/2 = .5	.50	.53 × .5 = 27
5	12	Mort	1	1/1 = 1,0	.00	.27 × 0 = 0
Grupa B						
1	2	Mort	5	1/5 = .2	.80	1 × .8 = .8
2	3	Mort	4	1/4 = .25	.75	.8 × .75 = .6
3	4	Mort	3	1/3 = .33	.67	.6 × .67 = .4
4	5	Mort	2	1/2 = .5	.50	.4 × .5 = .2
5	6	Surv	1	0/1 = .0	1.00	.2 × .1 = .2

În primul rând, pacienții sunt așezați în ordinea orei morții lor sau a ultimei viziuni în care au fost văzuți în viață. Să trecem prin al treilea rând pentru grupa A, ca exemplu.

Al treilea pacient a murit la 9 luni (coloanele 1 și 2). La începutul lunii a 9-a, erau trei pacienți cu risc de deces (din totalul de cinci pacienți care au intrat în studiu). Acest lucru se datorează faptului că unul dintre cei cinci pacienți murise deja în luna a 4-a (cazul #1), iar unul a fost văzut ultima dată în viață în luna a 5-a (cazul #2) și, prin urmare, nu era disponibil pentru a fi observat. Dintre acești trei pacienți cu risc la începutul lunii a 9-a, unul a murit (cazul #3). Deci, probabilitatea de a muri în luna a 9-a este de 1/3 și o numim q_i, unde i în acest caz se referă la luna a 9-a. Prin urmare, proporția care supraviețuiește în luna a 9-a este $p_i = 1 — q_i = 1 — .33 = .67$.

Proporția cumulativă de supraviețuire înseamnă proporția de supraviețuire până în luna a 9-a. Pentru a supraviețui până în luna a 9-a, un pacient trebuia să fi supraviețuit până la sfârșitul lunii 8 și să fi supraviețuit în luna 9. Astfel, este egală cu probabilitatea cumulată de a supraviețui până în luna a 9-a, care este 0,8, din coloana. 7 rândul 2 și supraviețuind în luna a 9-a, care este .67. Înmulțim aceste probabilități pentru a obține .8 × .67 = .53 ca probabilitate de supraviețuire până în luna a 9-a. Dacă reprezentăm aceste puncte ca în Figura 4.4 , observăm că cele două curbe de supraviețuire arată destul de diferit și că grupul A s-a descurcat mult mai bine.

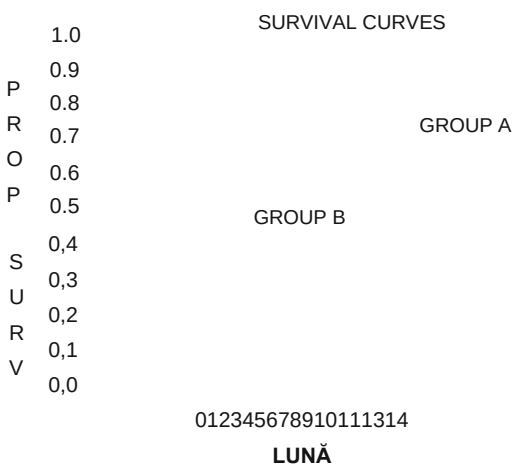


Figura 4.4 Curbele de supraviețuire

Analiza supraviețuirii devine mai complicată atunci când presupunem că pacienții care au fost pierduți de urmărire într-un anumit interval de timp ar fi murit în același ritm ca acei pacienți despre care am avut informații. Alternativ, putem face calculele presupunând că toți au murit în intervalul în care au fost pierdute pentru urmărire sau că toți au supraviețuit în acel interval.

Analiza supraviețuirii se poate face și în timp ce se controlează variabilele confuze, utilizând modelul hazardelor proporționale Cox.

4.19 Modelul de pericole proporționale Cox

Modelul de riscuri proporționale Cox este o formă de analiză multivariată a supraviețuirii care poate controla alți factori. Variabila dependentă este timpul până la eveniment (sau timpul de supraviețuire), care ar putea fi moartea, atacul de cord sau orice alt eveniment de interes. Acest lucru este în contrast cu regresia logistică multiplă, unde variabila dependentă este o variabilă da sau nu.

Modelul de riscuri proporționale Cox este utilizat în mod corespunzător atunci când există perioade de urmărire diferite, deoarece unele persoane s-au retras din studiu sau nu pot fi contactate. Persoanele care se încadrează într-una dintre aceste categorii sunt considerate a avea observații „cenzurate”. Dacă evenimentul de interes este, să zicem, un accident vascular cerebral, atunci oamenii care au murit în timpul studiului din cauze accidentale ar fi, de asemenea, „cenzurați”, deoarece nu am putea ști dacă ar fi suferit sau nu un accident vascular cerebral, dacă ar fi trăit până la sfarsitul studiului.

Coeficienții din această analiză pot fi utilizați pentru a calcula o estimare a riscului relativ de eveniment, după controlul celorlalte covariabile din ecuație. Această estimare din modelele de pericole proporționale Cox este mai precis numită „raportul de risc” sau HR. Dacă evenimentul de interes este moartea, atunci este pericolul de a muri într-un moment în timp într-un grup față de pericolul de a muri în celălalt grup. Raportul dintre aceste două pericole este HR. Ipoteza riscurilor proporționale înseamnă că presupunem că raportul dintre aceste două pericole este același în timp.

Un exemplu de interpretare a rezultatelor unei astfel de analize este dat din Programul de hipertensiune sistolică la vârstnici (SHEP). Acesta a fost un studiu pe 4.736 de persoane de peste 60 de ani cu hipertensiune sistolica izolata (de exemplu, persoane cu tensiune arteriala sistolica mare si sange diastolic normal).

presiune) pentru a vedea dacă tratamentul cu un diuretic în doză mică și/sau beta-blocant ar reduce rata accidentelor vasculare cerebrale în comparație cu rata din grupul de control tratat cu placebo.

Un eșantion de imprimare parțială de computer a unei analize de regresie Cox din studiul SHEP este prezentat mai jos. Evenimentul de interes este accidentul vascular cerebral în grupul placebo.

Variabila independenta	Coeficientul beta	se	$e^{B^{eta}} = RR$
Rasă	- 0,1031	.26070	0,90
Sex (masculin)	0,1707	.19520	1.19
Vârstă	0,0598	.01405	1.06
Istoric de diabet	0,5322	.23970	1,70
Fumatul (de bază)	0,6214	.23900	1,86

Să ne uităm la istoria diabetului. HR = $e^{,532}{}_2$ = 1,70, care este logaritmul natural e ridicat la puterea specificată de coeficientul beta ; e = 2,7183. (Nu întrebați de ce.) Aceasta înseamnă că o persoană cu hipertensiune sistolică netratată, care are antecedente de diabet, are un risc de 1,7 ori mai mare de a avea un accident vascular cerebral decât o persoană cu aceleași alte caracteristici, dar fără diabet. Acest lucru poate fi, de asemenea, declarat ca un risc cu 70% mai mare. Limitele de încredere de 95% pentru rapoartele de risc sunt 1,06, 2,72, ceea ce înseamnă că suntem 95% încrezători că riscul relativ de accident vascular cerebral pentru cei cu antecedente de diabet se află în intervalul cuprins între 1,06 și 2,72. Formula pentru intervalul de încredere de 95 % pentru raportul hazardului este

$$\text{Limita 1} = e^{[b^{eta} - 1:96(S:E:)]}$$

$$\text{Limit2} = e^{[b^{eta} + 1:96(S:E:)]}$$

Dacă avem de-a face cu o variabilă continuă, cum ar fi vârsta, HR-ul este dat pentru o unitate sau creștere de vârstă de 1 an. Rata hazardului pe 5 ani de creștere a vârstei este

$$e^{5\ beta} = e^{5\times.0598} = 1.35$$

Există o creștere cu 34% a riscului de AVC în viitor pe o vârstă cu 5 ani mai mare la momentul inițial, controlând toate celelalte variabile din model. La

calculați intervalele de încredere pentru acest exemplu, trebuie de asemenea să înmulțiți se cu 5 (precum și să înmulțiți beta cu 5), astfel încât intervalele de încredere de 95 % ale HR sunt [1,18, 1,55] pentru o creștere de 5 ani a vârstei.

Exemplul de mai sus se referă la o relație liniară între vârstă și riscul de accident vascular cerebral. Cu toate acestea, unele relații nu sunt liniare, ci mai degrabă au o formă J sau U. Secțiunea 4.25 oferă informații despre explorarea unei relații de formă J sau U între o variabilă și rezultat.

4.20 de încredere suprapuse și semnificație statistică

Să presupunem că priviți riscul de accident vascular cerebral în două grupuri, unul fiind tratat cu un anumit medicament și celălalt cu un alt medicament. Obțineți un raport al pericolelor și un interval de încredere de 95 % pentru fiecare grup. Să presupunem în continuare că intervalele de încredere se suprapun. De exemplu, în grupul A, aveți un HR de 1,30 și un interval de încredere de 95 % de 1,20–1,40, iar în celălalt grup, intervalul de încredere este 1,45–1,55. Puteți trage concluzia că cele două grupuri sunt semnificativ diferite unul de celălalt în ceea ce privește riscul de accident vascular cerebral? Dacă intervalele de încredere nu se suprapun, atunci cele două grupuri sunt statistic semnificativ diferite unul de celălalt și puteți respinge ipoteza lipsei de diferență la nivelul de semnificație .05. Dacă acestea ar fi fost intervale de încredere de 99 %, ați putea respinge ipoteza nulă a lipsei diferențelor la nivelul de semnificație de .01. Cu toate acestea, dacă intervalele de încredere se suprapun, nu puteți fi sigur că nu există o diferență semnificativă.

4.21 Confuzie prin indicație

Un tip de confuzie care poate apărea în studiile observaționale atunci când vă uitați la efectele tratamentului medicamentos asupra evenimentelor viitoare este confuzie prin indicație. De exemplu, să presupunem că doriți să comparați efectele asupra bolilor de inimă ale diferitelor medicamente pentru hipertensiune arterială. Dumneavoastră determinați ce medicamente antihipertensive iau participanții la studiu la început

examinare și apoi îi urmărești înainte în timp pentru a vedea cine dezvoltă boli de

inimă. Problema poate fi că motivul pentru care oamenii luau diferite medicamente la momentul inițial a fost că aveau indicații diferite pentru ele și medicii le-au prescris medicamente adecvate acelor indicații. Astfel, persoanelor cu boală de rinichi li s-ar putea să li se prescrie un alt medicament pentru a controla tensiunea arterială crescută decât cei cu angină pectorală sau decât cei fără alte afecțiuni medicale și fiecare dintre aceste indicații poate fi legată diferit de rezultatul bolilor de inimă. Doar un studiu clinic, în care pacienții sunt repartizați aleatoriu pentru fiecare tratament medicamentos, poate răspunde cu adevărat la întrebarea despre diferitele efecte ale medicamentelor.

Cu toate acestea, există modalități de a minimiza confuzia prin indicație în studiile observaționale; o modalitate este de a exclude din analiză persoanele care au angină sau boli de rinichi din exemplul de mai sus. O altă metodă care se folosește este analiza propensității. [24]

4.22 Analiza înclinației

Ideea generală din spatele analizei înclinației este că preziceți cine este probabil să ia medicamentul din variabilele independente pe care le-ați măsurat și calculați un indice de „propensitate” pentru a lua medicamentul. Scorul de înclinație este probabilitatea ca un individ să aibă expunerea (luarea medicamentului) condiționată de setul său de covariabile. Probabilitatea de probabilitate poate fi utilizată în trei moduri diferite și nu există un consens cu privire la care este cel mai bun: (1) ajustare de regresie, (2) stratificare și (3) potrivire.

(1) Prima, numită ajustare de regresie , este să utilizați scorul de înclinație (sau probabilitatea de înclinare) ca covariabilă în modelele de regresie logistică sau Cox (descrise în Secțiunea 4.19) care analizează relația dintre variabila dvs. de interes și rezultatul de interes. , așa cum este descris mai jos. Înclinația (sau probabilitatea de expunere) este apoi introdusă ca variabilă independentă în ecuația finală multivariată, împreună cu un subset de variabile pentru care controlați. Datele fiecărei persoane includ apoi valorile covariatelor și scorul său de propensiune.

De exemplu, în exemplul hipertensiunii arteriale, dacă căutăm să vedem dacă un blocant al canalelor de calciu este asociat cu mortalitatea, dorim să ținem cont de faptul că, într-un studiu observațional, este mai probabil

ca oamenii să li se fi prescris un blocant al canalelor de calciu dacă a avut angină, de exemplu, și știm că angina este legată de mortalitate. Apoi am putea face următorii pași:

(a) Calculați o regresie logistică multiplă în care Y (variabila dependentă) = 1 dacă se consumă droguri, 0 în caz contrar
X_i (variabile independente) = vârsta, rasa/etnia, angina pectorală, IMC, tensiunea arterială sistolică și alte covariabile care ar putea influența prescrierea unui blocant al canalelor de calciu, cum ar fi regiunea țării, indici socioeconomici și așa mai departe. Ar trebui să introduceți în această logistică atâtea variabile de bază câte ați măsurat pentru a obține un model de cea mai bună potrivire care prezice expunerea.

(b) Calculați un scor de înclinație pentru fiecare persoană (probabilitatea de expunere pe baza regresiei dezvoltate mai sus).

(c) Calculați regresia care vă interesează cu adevărat care este să determinați asocierea blocanților canalelor de calciu cu mortalitatea după controlul potențialilor factori de confuzie, unde Z = 1, dacă eveniment mortal , 0 în caz contrar (variabilă dependentă)
X_i = scorul de înclinație, vârsta, rasa/etnia, plus unele dintre celelalte covariabile relevante care se aflau în ecuația originală a înclinației.

(2) A doua modalitate de a utiliza scorurile de propensiune se numește stratificare (sau subclasificare). În această metodă, probabilitățile de înclinație obținute din ecuația de regresie logistică sunt împărțite în chintile (sau decile), adică strate bazate pe chintile de propensiune. La obținerea ulterioară a hazardului dintr-o ecuație de regresie Cox (a se vedea secțiunea 4.19), vom introduce medicamentul de interes ca o variabilă independentă și covariatele pe care dorim să le controlăm și să le stratificăm pe chintila (sau decilul) de propensiune.

Aceasta înseamnă că funcția de pericol de bază poate varia între chintilele de înclinație. Ce înseamnă asta? Ei bine, în fiecare strat de chintilă, oamenii sunt mai mult sau mai puțin la fel de probabil să ia droguri (au tendințe similare). Raportul de risc (care este pericolul evenimentului din grupul care utilizează medicamentul împărțit la pericolul din grupul care

nu utilizează drogul, considerat uneori riscul relativ) ar putea fi diferit în grupul cel mai puțin probabil să utilizeze medicamentul (chintila inferioară a tendinței) decât raportul de risc la cei mai susceptibili de a utiliza drogul (cei din chintila cea mai înaltă). Stratificarea compară apoi oamenii din grupuri similare în ceea ce privește tendința lor de a utiliza medicamentul.

Iată un exemplu dintr-un studiu privind utilizarea antidepresivelor în cadrul Women 'S Health Initiative. [25] Cercetătorii au fost interesați de efectele medicamentelor antidepresive asupra riscului cardiovascular. Populația studiată a constat din 136.293 de femei aflate în postmenopauză care au fost urmărite în medie 6 ani. Ne vom uita la efectele unui anumit tip de medicament antidepresiv numit inhibitori selectivi ai recaptării serotoninei (ISRS) asupra bolilor coronariene și asupra accidentului vascular cerebral.

Există mulți factori care pot influența un medic care prescrie un antidepresiv, în general, și un anumit antidepresiv, în special. Un factor deosebit de important este severitatea depresiei. Alți factori pot avea de-a face cu vârsta, alte afecțiuni comorbide, regiunea țării (utilizarea antidepresivelor poate fi mai frecventă în anumite locații geografice) și numeroase alte variabile care ar putea fi factori de confuzie potențiali. Anchetatorii au vrut să știe dacă noua utilizare a antidepresivelor a avut un efect asupra bolilor cardiace ulterioare sau riscului de accident vascular cerebral. Pentru a aborda potențiala confuzie în funcție de indicație, cercetătorii au obținut un scor de înclinație dintr-un model de regresie logistică pentru a prezice orice utilizare nouă a antidepresivelor din 33 de variabile demografice, stil de viață, factor de risc și comorbiditate măsurate la înscriere. (Unele studii folosesc mai mult de 100 de variabile pentru a crea un scor de înclinație, dacă aceste variabile sunt disponibile.) Astfel, aceste scoruri de înclinație au fost un compus ponderat al covariatelor individuale pentru fiecare persoană.

Pentru a determina cât de bună a fost regresia logistică la discriminarea între cei care au folosit antidepresive față de cei care nu au făcut - o, putem utilizați statistica c. Statistica c este o măsură a cât de mult mai bună decât șansa predicției noastre privind consumul de droguri se bazează pe variabilele din modelul nostru de regresie logistică. O statistică c de 0,5 ar însemna că modelul nostru logistic nu este mai bun decât șansa de a prezice apartenența la grupul de droguri față de non-aplic. O statistică c de 1,0 ar însemna că regresia logistică este perfect capabilă să discrimineze între persoanele care folosesc antidepresive față

de cei care nu le folosesc . În exemplul nostru, statistica c a fost 0,72, indicând o capacitate moderată a variabilelor incluse în model de a discrimina noua utilizare a antidepresivelor. (O statistică c de .7 sau mai mare este considerată rezonabilă, iar una de .8 sau mai mare este considerată a avea o puternică capacitate de discriminare.)

Scorurile de propensiune au fost apoi împărțite în grupuri de decile (quintilele sunt mai frecvent utilizate și sunt perfect adecvate). Modelul de regresie Cox a fost apoi rulat folosind instrucțiunea STRATA în software-ul SAS, versiunea 9.1; (SAS Institute Inc, Cary, Carolina de Nord). Variabila dependentă din regresia Cox a fost accidentul vascular cerebral, iar modelul Cox a fost stratificat în funcție de decilă a înclinației de a lua orice antidepresiv nou la începutul urmăririi și ajustat pentru următoarele covariabile: tensiunea arterială sistolica, indicele de masă corporală, măsura depresiei. pe o scară de depresie, utilizarea de hormoni, migrenă sau dureri de cap urâte, aspirina sau utilizarea antiinflamatoare nesteroidiene și antecedente de accident vascular cerebral sau atac de cord.

În primul rând, ne uităm la raportul de hazard neajustat pentru nimic. Pentru boala de inimă, care a fost definită ca un atac de cord sau un deces din cauza unei boli de inimă, HR neajustat a fost 1,28, intervalele de încredere de 95 % fiind 1,01–1,61 (de obicei scrise ca 1,28 (IC 95%: 1,01–1,61). Acest lucru ar duce noi să credem că noii utilizatori de antidepresive au avut un risc cu 28% mai mare de boli de inimă decât cei care nu folosesc intervalul nu se suprapune cu 1.0, suntem 95% încrezători că adevăratul HR se află în acel interval și, prin urmare, există un risc mai mare pentru utilizatori decât pentru neutilizatori. hazardul este de .95 (95% CI: .70–1.29), adică intervalul de încredere pe ture 1.0, iar concluzia noastră este că antidepresivul nu este asociat cu boli cardiace ulterioare.

Situația a fost diferită pentru accident vascular cerebral, unde HR neajustat a fost de 1,40 (95 % CI: 1,09–1,80) și propensiune HR ajustată a fost 1,45 (95 % CI: 1,08–1,97) – indicând că riscul de accident vascular cerebral a fost mai mare chiar și la utilizatori.

1.1 Biostatistică și epidemiologie: un primer pentru profesioniștii din domeniul sănătății după controlul pentru tendința și alte covariate. Prin urmare, este foarte important să se controleze factorii de confuzie.

Un al treilea mod de a folosi scorurile de propensiune este potrivirea. În această metodă, selectăm o persoană care utilizează medicamentul și găsim o persoană care nu ia medicamentul care are cel mai apropiat scor de înclinație; apoi, luăm următoarea persoană pe drog și găsim cea mai apropiată potrivire a acelei persoane care nu este pe drog și așa mai departe . Desigur, există programe de calculator care fac asta. Potrivirea se poate face până la 5 cifre, 4 cifre și 3 cifre. Asta trebuie să decidă cercetătorul.

Astfel, formăm două grupuri: tratați și de control (potrivit pe înclinație), și acesta este cel mai aproape de a simula randomizarea într-un studiu observațional. Putem apoi compara aceste două grupuri la ieșire . O problemă este că este posibil să nu putem găsi o potrivire pentru fiecare persoană tratată, caz în care dimensiunea eșantionului (grupurile potrivite) va fi mult mai mică decât am avea-o dacă am putea potrivi pe toată lumea și este posibil să nu avem suficientă putere pentru a detectează un efect (vezi capitolul despre putere). În exemplul nostru, cercetătorii au reușit să găsească doar 4.204 perechi potrivite sau 8.408 participanți, din cele 136.293 de femei care au fost în cohorta analitică inițială. Cele două grupuri potrivite au fost similare pe majoritatea caracteristicilor de bază, indicând că au fost bine echilibrate, cu excepția mai multor caracteristici care au fost controlate pentru modelele din Cox. Analiza potrivită a furnizat un raport de risc pentru accident vascular cerebral de 1,36 cu intervale de încredere de 95% de 0,88–2,10. Aceasta se compară cu 1,45 (IC 95%: 1,08–1,97) găsit cu ajustarea propensiunii prin stratificare. Analiza potrivită indică în continuare riscul crescut de consum de droguri, dar intervalul de încredere se suprapune 1,0, reflectând puterea inadecvată din cauza dimensiunii mai mici a eșantionului. Dar ordinul de mărime al raportului hazardului este aproximativ același.

În exemplul nostru de antidepresive și boli cardiovasculare, este dificil să distingem efectele antidepresivelor de efectele severității depresiei, deoarece depresia în sine este un factor de risc pentru bolile cardiovasculare. Din cauza problemelor de măsurare a depresiei și a variației tiparelor de prescripție, chiar și analiza înclinației poate să nu fie suficientă pentru a dezlega aceste două variabile.

Observați câteva lucruri despre scorurile de înclinație. Aceștia pot controla doar potențialele factori de confuzie cunoscuți, adică variabilele care sunt disponibile

Studiul. Desigur, factorii de confuzie necunoscuți nu pot fi controlați. Atunci când randomizați oamenii într-un grup de tratament și control într-un studiu clinic, este de așteptat ca randomizarea să echilibreze atât factorii de confuzie cunoscuți, cât și necunoscuți din cele două grupuri și, astfel, să evite părtinirea. Și astfel, cea mai bună modalitate de a determina efectele antidepresivelor asupra rezultatelor ar fi prin studii clinice controlate, randomizate, dublu-orb. Acest lucru prezintă multe provocări, inclusiv costul, timpul și fezabilitatea înscrierii unui eșantion suficient de mare pentru a avea suficientă putere pentru a detecta un efect. Între timp, este posibil să trebuiască să ne bazăm pe datele studiului observațional, dar ar trebui să ne amintim că un studiu clinic randomizat este considerat un „standard de aur" al dovezii de cauzalitate. (Vezi capitolul 6 în studiile clinice.)

4.23 Selectarea variabilelor pentru modelele multivariate

Să presupunem că dorim să determinăm efectul depresiei asupra evenimentelor ulterioare de boli de inimă folosind date dintr-un studiu prospectiv de urmărire. Putem rula modele de riscuri proporționale Cox pentru a obține riscul relativ de depresie pentru efectele bolilor de inimă, dar dorim să controlăm factorii de confuzie. În caz contrar, orice asociere pe care o vedem ar putea fi într-adevăr o reflectare a unei alte variabile, cum ar fi, să zicem, fumatul, care este legat de depresie și este, de asemenea, un factor de risc pentru boli de inimă. Cum vom decide ce variabile să introducem în model? Nu există un răspuns unic la această întrebare și experți diferiți au opinii oarecum diferite, deși este în general de acord că ar trebui incluși factorii de confuzie cunoscuți. Deci, am include variabile care sunt semnificativ legate de depresie și, de asemenea, de boli de inimă în rândul celor neexpuși, adică nedepresivi.

Nu am include variabile în model despre care se presupune că din experiența anterioară sunt fie foarte corelate cu depresia (denumită coliniară), fie intermediare în calea care leagă depresia cu bolile de inimă, cum ar fi un biomarker din sânge legat de bolile de inimă, care este crescut de depresie. Într-un astfel de caz, creșterea biomarkerului sanguin este intermediară între depresie și boli de inimă; poate fi prima manifestare a bolii cardiace. Nu ar trebui să ne ajustăm pentru variabile intermediare între expunere și rezultat.

Dacă includem variabile legate de expunere, dar nu de rezultat, aceasta va slăbi asocierea dintre expunere și rezultat. Dacă includem variabile legate de

rezultat, dar nu de expunere, putem crește precizia estimării relației expunere-rezultat. Dacă controlați variabilele legate de rezultat, dar nu de expunere, în general, se reduce variația inexplicabilă a variabilei rezultat și, prin urmare, se reduce varianța estimării parametrului pentru expunere, indiferent dacă este vorba de un hazard ratio, odds ratio , sau coeficientul într-o regresie liniară. Ideea este că trebuie folosită multă raționament în selectarea variabilelor pentru includere.

Obiectivul este de a vedea dacă efectele depresiei care au fost găsite rămân după luarea în considerare a altor factori de risc stabiliți. O strategie este să începeți prin a obține doar raportul de risc al depresiei și apoi să adăugați succesiv, unul câte unul, alți factori de confuzie potențiali pentru a vedea dacă modifică raportul de risc pentru depresie cu 10 % sau mai mult (deși acesta este un procent arbitrar). Variabilele care se califică după acest criteriu sunt păstrate în model. De exemplu, în studiul depresiei și deceselor din cauze cardiovasculare, în rândul femeilor aflate în postmenopauză înscrise în Women ' s Health Initiative, care nu au avut nicio boală cardiovasculară anterioară, riscul asociat cu controlul depresiei pentru vârstă și rasă a fost de 1,58; Adăugarea educației și a veniturilor la aceasta a dus la un hazard ratio de 1,52. Adăugarea de variabile suplimentare la model (diabet, hipertensiune arterială, fumat, colesterol ridicat care necesită pastile, utilizarea hormonilor, indicele de masă corporală și activitatea fizică) nu a schimbat lucrurile, rezultând un raport de risc de 1,50. Deci, s-a ajuns la concluzia că depresia este un factor de risc independent pentru decesul cardiovascular.

Acum, dacă cineva ar fi interesat să dezvolte un model care să prezică riscul (mai degrabă decât unul care să evalueze dacă un anumit factor de risc a contribuit independent la risc, ca în exemplul de mai sus), ar putea alege alte strategii, cum ar fi regresia în trepte. Regresia în trepte poate fi înainte sau înapoi, iar programele computerizate care calculează regresiile vă cer să specificați ce doriți.

Principiul de bază este că, în regresia progresivă înainte, introduceți mai întâi variabila care are cea mai mare corelație cu rezultatul dvs. , apoi continuați să adăugați variabile pe rând până când adăugați una care nu este semnificativă statistic la un nivel prestabilit, și apoi opriți. În
regresia în trepte înapoi, începeți cu toate variabilele potențiale care pot fi explicative și le renunțați una câte una, eliminând mai întâi pe cea mai puțin semnificativă (are cea mai mare valoare p), până când renunțarea la următoarea variabilă ar duce la o regresie mai slabă. model.

Mulți oameni nu le place regresia în trepte, deoarece este oarecum arbitrară; depinde de nivelurile de semnificație pe care le-ați ales să introduceți sau să

părăsiți modelul și, de asemenea, o variabilă poate avea un efect destul de diferit dacă este într-un model cu alte variabile care l-ar putea modifica, mai degrabă decât atunci când este doar în model. O altă strategie este să se uite la toate regresiile posibile, adică să se uite la toate cele două modele de variabile, apoi la toate cele trei modele de variabile posibile și așa mai departe. Îl selectați pe cel mai bun în funcție de cât de mult din variația variabilei dependente este explicată de model. O discuție excelentă despre selecția variabilelor în modelele epidemiologice este de Sander Greenland [26] și, de asemenea, în textele avansate notate în secțiunea Lecturi sugerate.

4.24 Interacțiuni: modele aditive și multiplicative

O interacțiune între două variabile înseamnă că efectul unei variabile asupra rezultatului de interes este diferit în funcție de nivelul celeilalte variabile, așa cum este descris în secțiunea 3,26 . Interacțiunile pot fi aditive , când efectul comun al două variabile este mai mare decât suma efectelor lor individuale, sau multiplicative, în cazul în care efectul comun al celor două variabile este mai mare decât produsul efectelor individuale ale fiecărei variabile.

Modelele de regresie logistică și Cox sunt în mod inerent multiplicative. Când spunem că fumatul prezintă un risc relativ de 2 pentru boala coronariană, de exemplu, ne referim la faptul că fumătorii au șanse de două ori mai mari decât nefumătorii de a face boala. Poate dorim să știm dacă există o interacțiune între fumat și hipertensiune arterială. Cu alte cuvinte, vrem să știm dacă fumatul printre hipertensivi are un efect mai mare asupra atacurilor de cord decât ne-am aștepta din cunoașterea riscurilor separate ale fumatului și ale hipertensiunii. Putem testa ipoteza lipsei de interacțiune față de ipoteza alternativă a unei interacțiuni, dar mai întâi trebuie să știm la ce ne-am aștepta de la un model multiplicativ dacă nu ar exista interacțiune.

Luați în considerare două variabile dihotomice A (fumatul) și B (hipertensiunea arterială).

Tabelul de mai jos arată modelul riscurilor relative așteptate în cadrul modelului multiplicativ dacă într-adevăr nu există interacțiune multiplicativă. Grupul de referință este $RR_{no,no} = 1$. Cu alte cuvinte, toate comparațiile noastre se referă la riscul dintre cei care nu au nici A, nici B, adică nefumătorii și

nehipertensivii în exemplul nostru. Rețineți că $RR_{da,\,da} = RR_{da,\,nu} \times RR_{nu,\,da} = 2,0 \times 1,5 = 3,0$. ($RR_{da,\,nu}$ este riscul relativ al lui B în absența lui A și $RR_{nu,\,da}$ este riscul relativ al lui A în absența lui B).

Risc relativ (RR)

	A nu	A da
B nr	$RR_{nu,\,nu} = 1$	$RR_{nu,\,da} = 1,5$
B da	$RR_{da,nu} = 2,0$	$RR_{da,\,da} = 3,0$

Dacă RR observat $_{da,\,da}$ este semnificativ diferit de 3.0, putem respinge ipoteza nulă a lipsei interacțiunii și concluziona că există o interacțiune multiplicativă.

Pentru a testa acest lucru statistic, am include un termen de produs în modelul nostru de regresie sau modelul hazardelor proporționale Cox (înmulțind valoarea variabilei B cu valoarea variabilei A pentru fiecare persoană pentru a obține o nouă variabilă care este produsul dintre A și B.), apoi calculați următoarea cantitate:

$$\frac{\beta \textit{ (coeficientul termenului de produs în regresia logistică)}}{\textit{eroare standard a } \beta}$$

Această cantitate pătrat este distribuită aproximativ ca chi-pătrat cu 1 grad de libertate.

Ceea ce testăm cu adevărat este dacă coeficientul β este semnificativ diferit de 0. Dacă este, atunci acest lucru este echivalent cu concluzia că $RR_{da,\,da}$ este semnificativ diferit de valoarea noastră așteptată de 3,0

Rețineți că, dacă există o interacțiune semnificativă, nu putem interpreta efectele principale ale aceluiași model în modul în care o facem de obicei. De exemplu, dacă a existat o interacțiune semnificativă între fumat și hipertensiune arterială,

ar trebui să ne stratificăm în funcție de hipertensiune, adică să privim efectele fumatului separat între hipertensivi și nehipertensivi.

Să presupunem că dorim să vedem ce fel de cifre de incidență ar putea da naștere tabelului de mai sus. Amintiți-vă, incidența este risc absolut, în timp ce riscul relativ este riscul absolut într-un grup în raport cu riscul absolut în grupul de referință. Cele două tabele de mai jos conțin ambele cifre de incidență care ar da naștere tabelului RR de mai sus, așa că puteți vedea că este posibil să aveți rate de incidență diferite sau riscuri care au același risc relativ.

Risk or incidence per 1,000

	A no	A yes
B no	$I_{no,no} = 20$	$I_{no,yes} = 30$
B yes	$I_{yes,no} = 40$	$I_{yes,yes} = 60$

	A no	A yes
B no	$I_{no,no} = 10$	$I_{no,yes} = 15$
B yes	$I_{yes,no} = 20$	$I_{yes,yes} = 30$

Riscul aditiv este mai rar testat, deși unii oameni cred că ar trebui să fie. Se calculează din diferența de riscuri absolute (mai degrabă decât din raportul riscurilor absolute). Sub ipoteza lipsei de interacțiune într-un model aditiv, ne-am aștepta la datele din tabelele de mai jos.

Incidence per 1,000

	A no	A yes
B no	$I_{no,no} = 20$	$I_{no,yes} = 30$
B yes	$I_{yes,no} = 40$	$I_{yes,yes} = 50$

Notă: Incidența $_{da\,,\,da}$ = incidența de bază + efectul A+ efectul lui B sau
$I_{yes,\,yes}$ — $_{ino,no}$ ^ (Ino,yes ^^^Ino,no) f ($_{yes,\,no}$ - Ino,no) — 20 + 10 + 20 — 50

Risk differences or attributable risk (AR) Per 1,000

	A no	A yes
B no	$AR_{no,no} = 0$	$AR_{no,yes} = 10$
B yes	$AR_{yes,no} = 20$	$AR_{yes,yes} = 30$

AR $_{da\,,\,da}$ — efectul lui A plus efectul lui B — 10 + 20 — 30

Dacă AR $_{da,da}$ este suficient de diferit de valoarea așteptată de 30, atunci putem concluziona că există o interacțiune pe scara aditivă.

Tabelul de risc relativ care corespunde tabelului de incidență pentru exemplul dat mai sus al modelului aditiv este:

Relative risk (RR)

	A no	A yes
B no	$RR_{no,no} = 1$	$RR_{no,yes} = 1.5$
B yes	$RR_{yes,no} = 2.0$	$RR_{yes,yes} = 2.5$

Astfel, valoarea așteptată a lui RR $_{da,\,da}$ în ipoteza nulă a interacțiunii fără aditivi este RR $_{da\,,\,da}$ — RR $_{da\,,\,nu}$ + RR $_{nu\,,\,da}$ — 1. ıf RR $_{da\,,\,da}$ este semnificativ diferită de așteptarea de mai sus, am putea respinge ipoteza nulă a riscului aditiv.

Interacțiunile depind de scară – adică dacă vorbim despre riscuri relative (multiplicative) sau riscuri atribuibile (aditive). Este înțelept să consultați un statistician pentru interpretări adecvate.

Rezumat

Efect de interacțiune a modelului aditiv
Vedeți dacă valoarea observată a $AR_{da,\,da}$ diferă de valoarea așteptată

$$AR_{yes,\,yes} - AR_{yes,\,no} + AR_{no,\,yes}$$
$$\text{or, } RR_{yes,\,yes} - RR_{yes,\,no} + RR_{no,\,yes} - 1$$

Efect de interacțiune model multiplicativ
Vedeți dacă valoarea observată a $RR_{da,\,da}$ diferă de valoarea așteptată

$$RR_{yes,\,yes} - RR_{yes,\,no} \times RR_{no,\,yes}$$

4.25 Relații neliniare: formă J sau formă U

În discuțiile despre regresia logistică și modelele de hazarde proporționale Cox, am vorbit despre relațiile liniare dintre covariatele din model și rezultatul sau variabila dependentă. Cu toate acestea, există multe fenomene biologice care au o relație în formă de J sau U în care riscul rezultatului este mai mare atât la nivelul inferior, cât și la cel mai înalt al variabilei de expunere și cel mai scăzut în intervalul mediu. Acest lucru este ilustrat de Figura 4.5 de mai jos, care arată relația dintre IMC (indicele de masă corporală

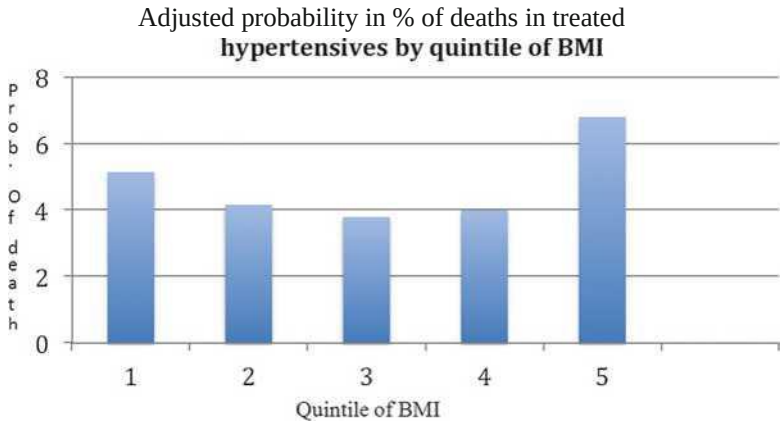


Figure 4.5 Adjusted probability in % of deaths in treated hypertensives by quintile of BMI

care este greutatea în kilograme împărțită la pătratul înălțimii în metri) până la moarte în rândul bărbaților și femeilor în vârstă tratați pentru hipertensiune sistolică în Programul de hipertensiune sistolică la vârstnici (SHEP). [27] Forma J sau U înseamnă că pe măsură ce IMC crește, rata mortalității crește, dar rata mortalității crește și la valori foarte scăzute ale IMC, de unde forma J sau U. Acest lucru se poate datora unei boli preexistente: oamenii care sunt foarte slabi pot fi slabit pentru ca sunt deja bolnavi si, desigur, vor avea mai multe sanse sa moara. Sau poate fi din cauza consecințelor fiziologice ale greutății foarte mici.

După cum puteți vedea, rata mortalității la 100, ajustată pentru covariate, este mai mare în chintila cea mai mică a IMC și cea mai mare chintilă și mai mică în chintila a 2-a, a 3-a și a 4-a.

Pentru a testa dacă există o relație neliniară semnificativă, adăugăm termenul pătratic (pătrat) la modelul de riscuri proporționale Cox, astfel încât am introduce atât IMC cât și IMC ^{2}. (Desigur, mai întâi ar trebui să calculăm IMC ^{2} pentru fiecare persoană). În exemplul nostru din acest studiu, coeficienții pe care i-am obținut, după controlul sau ajustarea pentru mai multe variabile, au fost pentru deces:

	Coeficientul	Eroare standard valorii coeficientului	p

IMC	- .3257	.1229	.008
IMC ^{2}	.005857	.0020	.003
	Pentru accident vascular cerebral		
IMC	- .2812	.1482	.06
IMC ^{2}	.0048	.0024	.05

Deoarece coeficientul pentru IMC ^{2} este semnificativ atât pentru deces, cât și pentru accident vascular cerebral, concluzionăm că există o relație neliniară între IMC și aceste rezultate. Nu putem interpreta coeficienții pentru IMC în mod obișnuit, deoarece aceștia reflectă relații liniare, deci nu au sensul obișnuit acum că știm că relația nu este liniară.

Din modelul Cox, putem calcula apoi raportul de risc sau riscul relativ (le folosim aici în mod interschimbabil) de deces și accident vascular cerebral, pentru diferite valori ale IMC în raport cu cea mai mică valoare a IMC (nadir-ul curbelor noastre) și vom obține diagramele din Figura 4.6 de mai jos. Calculul nadirului și riscurile relativ la nadir sunt prezentate sub figură.

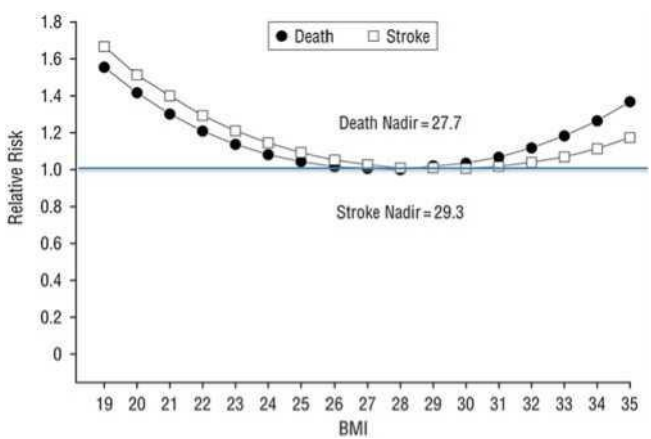


Figura 4.6 Relația dintre masa corporală scăzută și deces și accident vascular cerebral în programul hipertensiunii sistolice la vârstnici (Arch Intern Med. 2000; 160(4):494– 500. doi:10.1001/archinte.160.4.494)

Riscul relativ ajustat de deces și accident vascular cerebral letal și non-fatal în cadrul grupului de tratament activ în funcție de indicele de masă corporală (IMC) (calculat ca greutate în kilograme împărțită la pătratul înălțimii în metri). Variabilele au inclus nivelul educațional, istoricul de diabet, antecedentele de

infarct miocardic, istoricul de accident vascular cerebral, vârsta, nivelul de activitate, sexul, rasa, statutul de fumat, nivelul de colesterol, IMC, IMC ^{2}, tensiunea arterială sistolică (TAS) și TAS ^{2} (depinde de timp)

4.25 un nadir al relației cuadratice

Dacă coeficientul pătratic (termen pătrat) este semnificativ, dorim să știm la ce nivel al IMC este riscul relativ cel mai scăzut, care este nadir-ul curbei. Nadirul pentru accident vascular cerebral este calculat ca

> Nadir = — ½ (coeficient liniar / coeficient patratic) Nadir $_{\text{cursă}}$ = — ½ (— . 2812 /. 0048) = — . 5 × — 58 . 5833 = 29 . 3 Nadir $_{\text{moarte}}$ = — ½ (— 0 . 3257 / 0 . 005857) = 27 . 7

Aceasta înseamnă că cel mai scăzut risc de mortalitate pentru acest grup a avut loc la un IMC de 27,7, iar cel mai scăzut risc de accident vascular cerebral a avut loc la un IMC de 29,3.

4.26 b Riscul relativ la Nadir

De exemplu, dacă vrem să vedem care este riscul unui IMC de 20 în raport cu riscul de IMC de 29,3 (nadir) pentru accident vascular cerebral, îl putem calcula așa cum se arată mai jos.

RR = e^{k} (Secțiunile 4.17 și 4.19)

k = coeficient liniar (IMC_i - IMC_{nadir})

+ coeficientul patratic (IMC_i^2 — IMC^2_{nadir}) k = — : 2812 (20 — 29 : 3) + : 0048 (20^2 — $29 : 3^2$)

= — : 2812 (— 9 : 3) + : 0048 (— 458 : 49)

= 2 : 6152 — 2 : 2008 = : 4144

$e^{k} = e^{:4144}$ = 1 : 51

Astfel, comparativ cu un IMC de 29,3, a avea un IMC de 20 este asociat cu un risc cu 51% mai mare de accident vascular cerebral după ajustarea pentru mai multe covariate. Pentru a obține punctele pentru Figura 4.6 , facem astfel de calcule pentru fiecare valoare IMC pentru accident vascular cerebral și deces separat, fiecare cu propriul nadir.

4.25c Risc relativ Comparând oricare două valori

Pentru a calcula riscul relativ de deces pentru o anumită valoare a IMC în comparație cu orice altă valoare a IMC,

k = coeficientul liniar al termenului × (IMC_i — IMC_2) + coeficientul termenului pătrat (IMC_i^2 — IMC_2^2)

Notă : dacă coeficientul termenului pătrat nu a fost semnificativ, atunci k = coeficient liniar [IMC_i — IMC_2]

Chapter 5
MAI MULT DESPRE PROIECTARE

Mai degrabă aș profita de șansa mea ca unii trădători să scape de depistare decât să răspândească în străinătate un spirit de suspiciune și neîncredere generală, care acceptă zvonurile și bârfele în locul unei anchete neconsternate și neintimidate.

Judecătorul Learned Hand
octombrie 1952

5.1 Sensibilitate, oraș specific și concepte înrudite

Problema în utilizarea testelor de screening sau de diagnosticare este de a găsi un compromis adecvat între dorința de a detecta boala la persoanele care o au cu adevărat și dorința de a evita să credeți că ați detectat-o la persoanele care cu adevărat nu o au . .

O modalitate importantă de a vizualiza testele de diagnostic și screening este prin analiza sensibilității. Definițiile termenilor și simbolurilor relevanți sunt următoarele:

T+ înseamnă test pozitiv , T – înseamnă test negativ , D+ înseamnă boală, D – înseamnă că nu ai boală. Simbolul | înseamnă, „având în vedere că”, astfel încât P(T + | D —) înseamnă test pozitiv , având în vedere că nu există boală sau D — .

Sensibilitate: proporția de persoane bolnave pe care testul le clasifică drept pozitive,

$$= —= \frac{a}{P(T+|D+)}; \quad a + c$$ (probabilitatea testului pozitiv , dată fiind boala)

Specificitate: proporția de persoane neboli pe care testul o clasifică drept negativ,

$$= b + d = \frac{d}{P(T—|D—)};$$ (probabilitatea testului negativ, dat fiind nicio boală)

Rata fals-pozitive: proporția de persoane neboli pe care testul o clasifică (incorect) drept pozitivă,

b
-- ---- - = P (T + ı D —) ; (probabilitateatestului pozitiv, dat fără boală) *b* + *d*

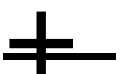
Rata fals-negative: proporția de persoane bolnave pe care testul le clasifică (incorect) drept negative,

c
------- = P (T — ı D +); (probabilitatea unui test negativ, dată de boală) *a* + *c*

Valoarea predictivă a unui test pozitiv: proporția de teste pozitive care identifică persoanele bolnave,

o
----- = P (D + ı T +); (probabilitate de boală dat pozitivtest) a + b

Valoarea predictivă a unui test negativ: proporția de teste negative care identifică corect persoanele care nu sunt bolnave,



d = P (D — ı T —) ; (probabilitatea de a nu se da boalatest negativ)

Precizia testului: proporția tuturor testelor care sunt clasificări corecte,

$$\frac{a + d}{a + b + c + d}$$

Raportul de probabilitate a testului pozitiv: raportul dintre probabilitatea unui test pozitiv, având în vedere boala, și probabilitatea unui test pozitiv, dacă nu există boală,

P (T + ı D +)
P ı T _|_ ı D —) = pozitivtest, dat boalăversuspozitivtest, dat fără boală
sensitivity sensitivity
fals pozitive 1 — specificitate

Raportul de probabilitate a unui test negativ:

$$\frac{\mathbf{P}(T - \mid D+)}{\mathbf{P}(T - \mid D-)}$$ = test negativ, dată boalăversus test negativ , dată fără boală

1 — specificitatea

sensibilitate

Rețineți și următoarele relații:

(1) Specificitatea + rata fals pozitive = 1:

$$\frac{db}{b + d^{+} b + d}$$

prin urmare, dacă specificitatea unei test is increased, the false-rate pozitive este scăzută.

(2) Sensibilitate + rata fals-negative = 1:

$$\frac{ac}{a + c\ a + c} \qquad 1$$

prin urmare, dacă sensibilitatea unui test este crescută, rata fals negative va fi scăzută.

Probabilitatea pretestare a bolii : Probabilitatea pretestare a unei boli este prevalența acesteia. Neștiind nimic despre un individ și în absența unui test de diagnostic, cea mai bună presupunere a probabilității ca pacientul să aibă boala este prevalența bolii.

Probabilitatea de boală posttest : După obținerea rezultatelor testului, probabilitatea posttest de boală dacă testul este normal este *c/* (*c* + *d*), iar dacă este anormală, probabilitatea posttest este *a/* (*a* + *b*). Ultima este aceeași cu *valoarea predictivă a unui test pozitiv.*

Un test de diagnostic bun este unul care vă îmbunătățește estimarea despre starea bolii pacientului față de estimarea pe care ați face-o doar pe baza prevalenței generale a bolii. Cu toate acestea, de interes principal pentru un clinician este

valoarea predictivă a unui test pozitiv (PV+), care este proporția de persoane care au un test pozitiv și care au într-adevăr boala, a/(a +b), și *valoarea predictivă a unui test negativ (PV -),* care este proporția de persoane cu un test negativ care cu adevărat nu au boala, *d/* (*c + d*).

Sensibilitatea și specificitatea sunt caracteristici ale testului în sine, dar valorile predictive sunt foarte mult influențate de cât de frecventă este boala. De exemplu, pentru un test cu 95 % sensibilitate și 95 % specificitate folosit pentru a diagnostica o boală care apare doar la 1 % dintre oameni (sau 100 din 10.000), am avea următoarele:

		Boală		
		Da +	Nu ^^^^^^	
Test	+	95	495	590
	^^^	5	9.405	9.410
		100	9.900	10.000

PV+ este 95/590 = .16; asta înseamnă că doar 16 % dintre toate persoanele cu rezultate pozitive la test au într-adevăr boala; 84% nu au boala, chiar dacă testul este pozitiv. PV - , totuși, este de 99,9 %, ceea ce înseamnă că dacă un pacient are un rezultat negativ al testului, poți fi aproape complet sigur că nu are cu adevărat boala . Valoarea practică a unui test de diagnostic depinde de o combinație de sensibilitate, specificitate și prevalență a bolii, toate acestea determinând valorile predictive ale rezultatelor testului.

Dacă prevalența bolii este mare, valoarea predictivă a unui test pozitiv va fi și ea mare, dar un test bun ar trebui să aibă o valoare predictivă mare, chiar dacă prevalența bolii este scăzută. Să aruncăm o privire asupra relației dintre prevalența bolii și sensibilitate, specificitatea și valoarea predictivă a unui test, prezentată în figura 5.1 .

Să luăm, de exemplu, un test care are o sensibilitate de .95 și o specificitate de .99. Aceasta înseamnă că acest test va eticheta corect ca fiind bolnavi 95% dintre indivizii cu boală și va eticheta corect ca nebolnași 99% dintre indivizii fără boală. Să luăm în considerare o boală a cărei prevalență este de 10%, adică 10% din populație are această boală, și să ne uităm acum și să vedem care este valoarea predictivă a unui pozitiv.

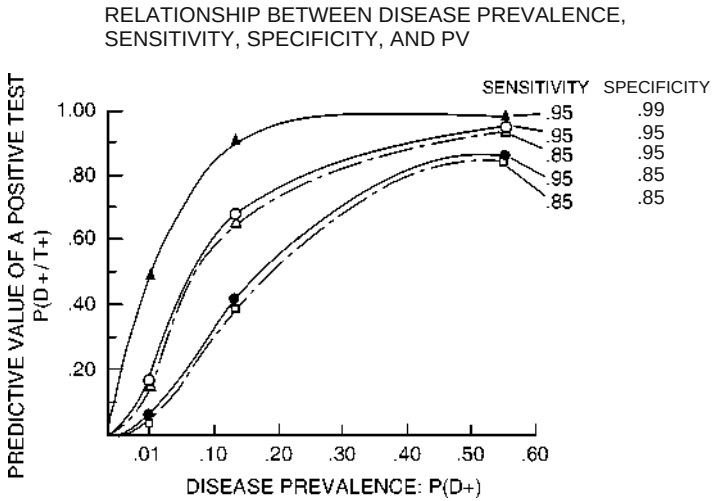


Figure 5.1 Relationship between prevalence, sensitivity, specificity and predictive value

testul este. Remarcăm că este de aproximativ .90, ceea ce înseamnă că 90% dintre indivizii cu un test pozitiv vor avea boala. Putem vedea că valoarea predictivă a unui test pozitiv scade la aproximativ .70 pentru un test care are o sensibilitate de .95 și o specificitate de .95 și putem vedea că mai scade la aproximativ .40 pentru un test care are o sensibilitate de .95 și o specificitate de .85. Cu alte cuvinte, doar 40% dintre persoanele cu un test pozitiv ar avea cu adevărat boala pentru un test care are acea sensibilitate și specificitate.

Un lucru pe care îl puteți observa imediat este că, pentru boala cu prevalență scăzută , valoarea predictivă a unui test pozitiv scade destul de brusc. Celălalt lucru pe care îl puteți observa aproape imediat este că diferența mare de sensibilitate face o mică diferență în valoarea predictivă a unui test pozitiv și că o mică diferență de specificitate face o mare diferență în valoarea predictivă a unui test pozitiv. Aceasta înseamnă că caracteristica unui test de screening descrisă prin specificitate este mai importantă în determinarea valorii predictive a unui test pozitiv decât sensibilitatea.

Figura 5.2 ne arată o situație a unui test care este practic perfect. Un test care are o sensibilitate de 0,99 și o specificitate de 0,99 este astfel încât, la cele mai multe niveluri de prevalență a bolii, probabilitatea de apariție a bolii, având în vedere un test normal sau negativ

SENSIBILITATE = .99

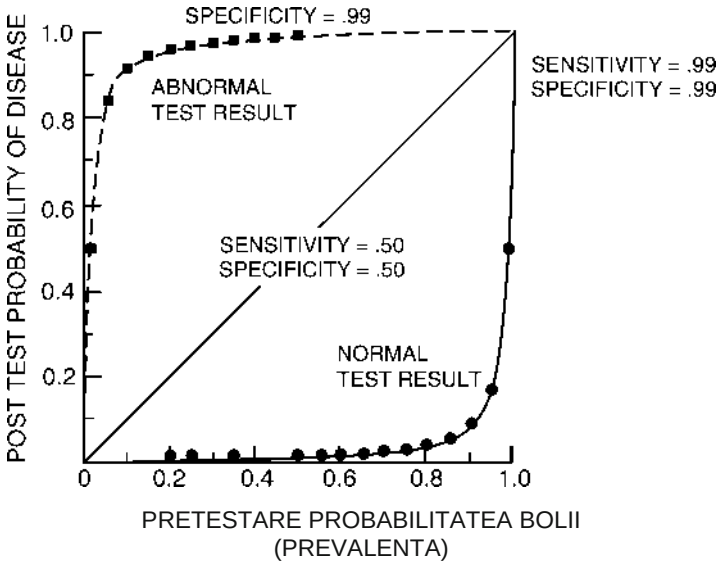


Figura 5.2 Probabilitatea post-test de apariție a bolii versus prevalență pentru diferite
sensibilități și specificități

rezultat, este foarte scăzut. Acesta ar fi un test foarte bun și, cu cât ne putem apropia de o astfel de situație, cu atât testul de diagnostic este mai bun. Linia diagonală din centru reprezintă un test cu o sensibilitate de .50 și o specificitate de .50, și acesta, desigur, este un test complet inutil pentru că se poate observa că la prevalența bolii de .4, probabilitatea a bolii dat un test pozitiv este, de asemenea, 0,4, care este aceeași cu probabilitatea bolii fără a face niciun test și aceasta se referă la fiecare nivel de prevalență. Prin urmare, un astfel de test este complet inutil, în timp ce un test cu sensibilitate și specificitate de .99 este excelent și orice lucru între ele reprezintă o utilitate diferită pentru teste. Aceasta, deci, este o modalitate analitică de a analiza procedurile de testare de diagnosticare.

Un exemplu deosebit de relevant al implicațiilor prevalenței asupra valorii predictive este cazul screening-ului pentru prezența infecției cu virusul SIDA. Deoarece această boală este în general fatală, tratabilă, dar incurabilă în prezent, provoacă anxietate mare, are un stigmat atașat și presupune costuri mari, nu ar dori să folosim o strategie de screening care etichetează în mod fals oamenii ca fiind pozitivi pentru HIV, SIDA. virus.

Să ne imaginăm că avem un test pentru acest virus care are o sensibilitate de 100% și o specificitate de 99,995%, clar un test foarte, foarte bun.

Să presupunem că o aplicăm în mod obișnuit tuturor donatoarelor de sânge de sex feminin, la care prevalența este estimată a fi foarte scăzută, să zicem 0,01%. În comparație, să presupunem că o aplicăm și bărbaților homosexuali din San Francisco, la care prevalența poate fi estimată la 50% (acest lucru este doar pentru scopuri ilustrative). Pentru fiecare 100.000 de astfel de persoane verificate, am avea valori așa cum se arată în tabelul de pe pagina următoare.

Deși în ambele grupuri ar fi identificați toți cei care au avut cu adevărat boala, în rândul donatorilor de sânge de sex feminin, o treime din toate persoanele care au fost testate pozitiv nu ar avea cu adevărat boala; printre homosexualii de sex masculin, doar 6 din 100.000 de persoane cu un test pozitiv ar fi etichetate fals.

Valoarea predictivă pozitivă ca funcție a prevalenței

Caracteristicile testului:

Sensibilitate = 100%; Specificitate = 99,995%; Rata fals pozitive = .005%

A. DONĂTORI DE SÂNGE FEMININ Starea adevărată **Prevalența = .01%,**

Rezultatul ecranului	**HIV: +**	**HIV: –**	
+	10	5	15
–	0	99.985	99.985
	10	99.990	100.000

PV+ = 10/15 = .66667

B. HOMOSEXUALI BĂRBAȚI **Prevalență = 50%,**

Adevărata stare

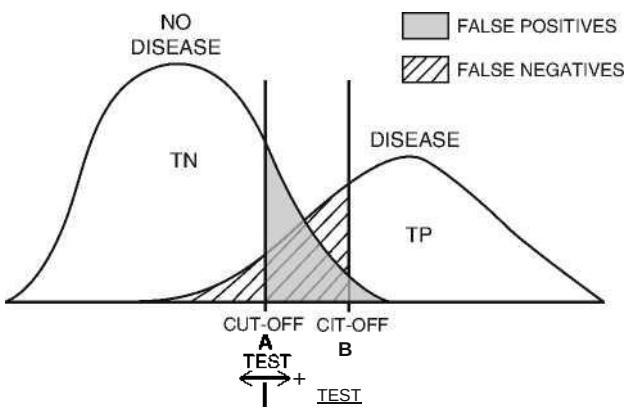
Rezultatul ecranului	**HIV: +**	**HIV: –**	
+	50.000	3	50.003
–	0	49.997	49.997
	50.000	50.000	100.000

PV+ = 50.000/50.003 = .99994

5.2 Punctul de tăiere și efectele acestuia asupra sensibilității și orașul specific

Am discutat despre sensibilitatea și specificitatea ca caracteristice ale unui test de diagnostic; cu toate acestea, ele pot fi modificate prin alegerea punctului de limită între normal și anormal. De exemplu, putem dori să diagnosticăm pacienții ca hipertensivi sau normotensivi după tensiunea arterială diastolică. Să spunem că oricine are o presiune diastolică de 90 mmHg sau mai mult va fi clasificat drept „hipertensiv". Deoarece tensiunea arterială este o caracteristică continuă și variabilă, la orice măsurătoare , un individ de obicei nehipertensiv poate avea o tensiune arterială diastolică de 90 mmHg sau mai mult și, în mod similar, un individ cu adevărat hipertensiv poate avea o singură măsură mai mică de 90 mmHg. Cu un punct de limită de 90 mmHg, vom clasifica unii indivizi nehipertensivi ca hipertensivi, iar aceștia vor fi fals pozitivi. De asemenea, vom eticheta unele persoane hipertensive ca fiind normotensive și acestea vor fi fals negative. Dacă am avea un punct de limită mai strict, de exemplu, 105 mmHg, am clasifica mai puțini nehipertensivi ca fiind hipertensivi, deoarece mai puțini indivizi normotensivi ar avea o citire atât de mare (și ar avea mai puține pozitive false).

Cu toate acestea, am avea mai multe fals negative (adică, mai mulți dintre oamenii noștri cu adevărat hipertensivi s-ar putea înregistra ca având tensiune arterială diastolică mai mică de 105 mmHg în orice ocazie). Aceste concepte sunt ilustrate în Figura 5.3 .



CUTOFF A:
SENSIBILITATE MAI MARE;
SPECIFICITATE INFERIORĂ;
MAI FALSE POZITIVE

CUTOFF B:
SENSIBILITATE MICĂ;
SPECIFICATĂ MAI MARE;
MAI MULTE FALSE NEGATIVE

Figura 5.3 Diferite puncte limită ale testului și fals pozitive și fals negative

Există două distribuții ale populației, cea bolnavă și cea nebolnică, și se suprapun pe măsura de interes, indiferent dacă este vorba despre tensiunea arterială, glicemia sau alte valori de laborator. Există foarte puține teste de screening care nu au nicio suprapunere între indivizii normali și cei bolnavi.

Un obiectiv în luarea deciziei asupra unui punct de limită este de a găsi echilibrul adecvat între fals pozitive și false negative. După cum puteți vedea în Figura 5.3 , când punctul de limită este la A, toate valorile din dreapta lui A sunt numite pozitive (se consideră că pacientul are boala). De fapt, totuși , pacientul cu o valoare la dreapta limită A ar putea proveni din populația de persoane nebolnave, întrucât o proporție de persoane perfect normale poate avea totuși valori mai mari decât cele de deasupra lui A, așa cum se vede în curba normală. Zona din dreapta lui A sub curba fără boală reprezintă fals pozitiv.

Dacă un individ are o valoare de test în stânga limitei A, el poate fi un negativ adevărat sau poate fi un negativ fals deoarece o proporție de indivizi cu boală poate avea totuși valori mai mici decât limitele A. Zona de sub „boală”. ” curba din stânga limitei A reprezintă proporția de fals negative.

Dacă mutam punctul de tăiere de la A la B, vedem că scădem aria din dreapta limitei, scăzând astfel numărul de fals pozitive, dar crescând numărul de fals negative. În mod corespunzător, cu cutoff-ul A, avem o probabilitate mai mare de a identifica corect bolnavii cu adevărat, adică de a obține mai multe pozitive adevărate, dând astfel testului cu cutoff A o sensibilitate mai mare. Cu cutoff-ul B, avem mai puține șanse să obținem adevăratele pozitive (sensibilitate mai scăzută), dar mai probabil să identificăm corect negativele adevărate (specificitate mai mare).

Astfel, prin deplasarea punctului de limită dincolo de ceea ce numim un test pozitiv, putem modifica caracteristicile de sensibilitate și specificitate ale testului. Alegerea limitei, cu excepția cazului în care există un motiv fiziologic special, se poate baza pe luarea în considerare a consecințelor relative ale existenței prea multor fals pozitive sau prea multe fals negative. Într-un test de screening pentru cancer, de exemplu, ar fi de dorit să existe un test de sensibilitate ridicată (și puține negative false), deoarece eșecul de a detecta această afecțiune din timp este adesea fatal. Într-un test de screening în masă pentru o afecțiune mai puțin gravă sau pentru una în care detectarea precoce nu este critică, ar putea fi mai de dorit să existe o specificitate ridicată pentru a nu suprasolicita sistemul de furnizare a asistenței medicale.

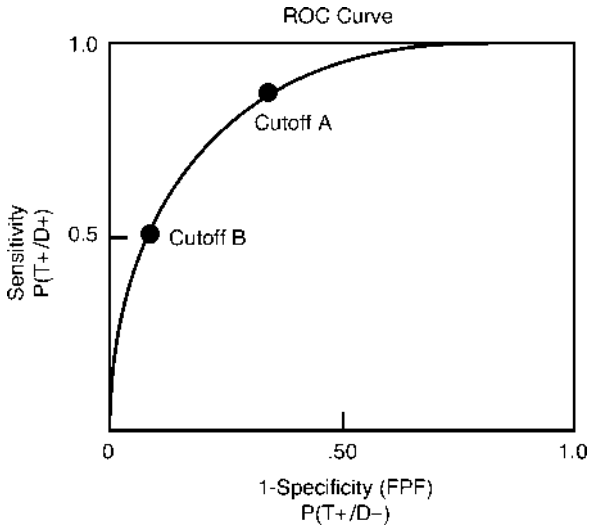


Figura 5.4 Curba caracteristicii de funcționare a receptorului (ROC), sensibilitate față de fracția fals pozitivă

cu prea multe false pozitive. Luarea în considerare a costurilor poate intra și în alegerea punctului limită.

Relația dintre sensibilitate (capacitatea de a identifica corect indivizii bolnavi) și fracțiile fals pozitive este prezentată în Figura 5.4 .

Aceasta se numește curba caracteristicii de funcționare a receptorului (ROC) a testului. Adesea putem selecta punctul de limită între normal și anormal, în funcție de compromisul pe care suntem dispuși să îl facem între sensibilitate și proporția de fals pozitive.

Putem vedea că cu limita A, în timp ce putem detecta un procent mai mare de indivizi cu adevărat bolnavi, vom avea și o proporție mai mare de rezultate fals pozitive, în timp ce cu limita B vom avea mai puține false pozitive , dar va fi mai puțin probabil. pentru a-i depista pe cei cu adevărat bolnavi. Testele de screening ar trebui să aibă trasate curbele ROC corespunzătoare.

Chapter 6
CELE MAI MULT DESPRE TRESTILE CLINICE

Nu este o sarcină ușoară să-ți îndrepti cuiva drumul de la adevăr la adevăr prin erori așezatoare.

Peter Marc Latham
1789–1875

Nu l- as fi vazut daca nu as crede !

Atribuit lui Yogi Berra

Din păcate, uneori oamenii de știință văd ceea ce cred în loc să creadă ceea ce văd. Studiile clinice randomizate, controlate sunt menite să evite acel și alte tipuri de părtinire.

Un studiu clinic randomizat este un experiment prospectiv pentru a compara una sau mai multe intervenții cu un grup de control pentru a determina eficacitatea intervențiilor. Un studiu clinic poate compara valoarea unui medicament față de un placebo. Un placebo este o substanță inertă care arată ca medicamentul testat. Se poate compara o nouă terapie cu o terapie standard în prezent, chirurgicală cu intervenție medicală, două metode de predare a lecturii și două metode de psihoterapie. Principiile se aplică oricărei situații în care problema cine este expus la care condiție se află sub controlul experimentatorului și metoda de atribuire este prin randomizare.

6.1 Caracteristicile studiilor clinice randomizate

(1) Există un grup de pacienți care sunt desemnați pacienți de studiu. Toate criteriile trebuie stabilite și îndeplinite înainte ca un potențial candidat să poată fi considerat eligibil pentru studiu. Orice excluderi trebuie specificate.
(2) Orice motiv pentru excluderea unui potențial pacient de la participarea la studiu trebuie specificat înainte de începerea studiului. În caz contrar, poate apărea părtinire neintenționată. De exemplu, să presupunem că ești

compararea chirurgiei de bypass coronarian cu utilizarea unui nou

medicament pentru tratamentul bolii coronariene. Să presupunem că vine un pacient care este eligibil pentru studiu și este repartizat la tratamentul chirurgical. Să presupunem că acum descoperi că pacientul are boală de rinichi. Decizi să-l excluzi din studiu pentru că crezi că ar putea să nu supraviețuiască intervenției chirurgicale cu rinichi afectați. Dacă ajungeți să excludeți sistematic toți pacienții bolnavi de la tratamentul chirurgical, puteți părăsi rezultatele în favoarea pacienților mai sănătoși, care au șanse mai mari de supraviețuire în orice caz. În acest exemplu, boala renală ar trebui să fie un criteriu de excludere aplicat pacienților înainte ca aceștia să fie alocați oricărui grup de tratament.

(3) Odată ce un pacient este eligibil, el sau ea este repartizat aleatoriu în grupul experimental sau de control. Atribuirea aleatorie nu este o atribuire „hap hazard", ci mai degrabă înseamnă că fiecare persoană are șanse egale de a fi un pacient experimental sau de control. De obicei, se realizează prin utilizarea unui tabel de numere aleatorii, descris mai târziu, sau prin numere aleatoare generate de computer.

(4) Studiile clinice pot fi dublu-orb, în care nici medicul curant, nici pacientul nu știu dacă pacientul primește tratament experimental sau placebo; acestea pot fi single-orb, în care medicul curant știe în ce grup se află pacientul, dar pacientul nu știe. Un studiu dublu-orb conține cea mai mică părtinire, dar uneori nu este posibil din motive etice sau practice. De exemplu, medicul poate avea nevoie să cunoască grupul căruia îi aparține pacientul, astfel încât medicația să poată fi ajustată pentru bunăstarea pacientului. Există, de asemenea, studii în care atât pacienții, cât și medicii cunosc grupul de tratament, ca în studiile care compară mastectomia radicală versus lumpectomia pentru tratamentul cancerului de sân. Când mortalitatea este rezultatul, posibila părtinire introdusă este minimă, cu condiția ca criteriile de excludere să fi fost specificate și aplicate înainte ca eligibilitatea să fie determinată în final și ca randomizarea participanților eligibili în grupurile de tratament să fi fost efectuată în mod corespunzător.

(5) În timp ce studiile clinice compară adesea un medicament sau un tratament cu placebo, ele pot compara, de asemenea, două tratamente între ele,

sau un tratament și „îngrijire obișnuită". În mod evident, studiile care compară o intervenție cu „îngrijirea obișnuită" nu pot fi orbite, de exemplu, comparând o intervenție nutrițională de slăbire cu dieta „obișnuită"; cu toate acestea, evaluarea efectului (măsurarea greutății sau

a tensiunii arteriale, sau un efect ipotetic al pierderii în greutate) ar trebui să se facă într-un mod orb, evaluatorul neștiind la ce grup a fost repartizat participantul.

(6) Este esențial ca grupul de control să fie cât mai asemănător cu grupul de tratament, astfel încât diferențele de rezultat să poată fi atribuite diferențelor de tratament și nu caracteristicilor diferite ale celor două grupuri. Randomizarea ajută la realizarea acestei comparabilități.

(7) Suntem preocupați aici de studiile de fază III. Noile medicamente trebuie să fie supuse unor studii de fază I și II, care determină toxicitatea și, respectiv, siguranța și eficacitatea. Aceste studii sunt efectuate pe un număr mic de voluntari. Studiile de fază III sunt studii clinice mari, suficient de mari pentru a oferi un răspuns la întrebarea dacă medicamentul testat este mai bun decât placebo sau decât un medicament de comparație.

6.2 Scopurile randomizării

Principiul de bază în proiectarea studiilor clinice sau a oricărei investigații științifice este evitarea părtinirii sistematice. Când nu se știe care variabile pot afecta rezultatul unui experiment, cel mai bun mod de a evita părtinirea sistematică este de a repartiza indivizii în grupuri aleatoriu. Randomizarea are scopul de a asigura o distribuție aproximativ egală a variabilelor între diferitele grupuri de indivizi care sunt studiate. De exemplu, dacă studiezi efectul unui medicament antidiabetic și știi că factorii de risc cardiac afectează mortalitatea în rândul diabeticilor, nu ai dori ca toți pacienții din grupul de control să aibă boli de inimă, deoarece asta ar influența în mod clar rezultatele. Prin repartizarea aleatorie a pacienților în grupul de medicament și grupul de control, vă puteți aștepta ca distribuția pacienților cu probleme cardiace să fie comparabilă în cele două grupuri. Deoarece există multe variabile care sunt necunoscute, dar care pot avea o influență asupra rezultatelor, randomizarea este o asigurare împotriva părtinirii necunoscute și neintenționate. Desigur, când

Tratând cu variabile cunoscute a fi relevante, acestea se pot lua în considerare prin stratificarea și apoi randomizarea în cadrul straturilor. De exemplu, vârsta este o variabilă relevantă pentru rezultatul diabetului. Pentru a stratifica în funcție de vârstă, puteți selecta patru grupe de vârstă pentru studiu: 35–44, 45–54, 55–64, 65

plus. Fiecare grup este considerat un strat. Atunci când un pacient intră în studiul clinic, stratul său de vârstă este mai întâi determinat și apoi este repartizat aleatoriu fie în grupul experimental, fie în grupul de control. Sexul este o altă variabilă care este adesea tratată prin stratificare.

Un alt scop al randomizării are de-a face cu faptul că tehnicile statistice utilizate pentru compararea rezultatelor între grupurile de pacienți studiate sunt valabile în baza anumitor ipoteze care decurg din randomizare . Motivele matematice ale acestui lucru pot fi găsite în textele mai avansate enumerate în lecturi sugerate.

Trebuie amintit că, uneori, randomizarea nu reușește să rezulte grupuri comparabile din cauza întâmplării. Acest lucru poate prezenta o problemă majoră în interpretarea rezultatelor, deoarece diferențele de rezultat pot reflecta diferențe în compoziția grupurilor în ceea ce privește caracteristicile de bază, mai degrabă decât efectul intervenției. Sunt disponibile metode statistice pentru a ajusta caracteristicile de referință despre care se știe că sunt legate de rezultat. Unele dintre aceste metode sunt regresia logistică, modelele de riscuri proporționale Cox și analizele de regresie multiplă.

6.3 Cum se efectuează o atribuire aleatorie

Alocarea aleatorie într-un grup experimental sau într-un grup de control înseamnă că fiecare individ eligibil are șanse egale de a fi în fiecare dintre cele două grupuri. Acest lucru este adesea realizat prin utilizarea tabelelor cu numere aleatorii. De exemplu, un extras dintr-un astfel de tabel este prezentat mai jos:

48461 70436 04282
76537 59584 69173

Utilizarea sa ar putea fi după cum urmează. Toate persoanele cu număr par sunt alocate grupului de tratament, iar toate persoanele cu număr impar sunt alocate grupului de control. Prima persoană care intră în studiu primește primul număr din

listă, următoarea persoană primește următorul număr și așa mai departe. Astfel, primei persoane i se atribuie numărul 48461, care este un număr impar și atribuie pacientul grupului de control. Următoarea persoană primește 76537; acesta este, de asemenea, un număr impar, deci și el/ea aparține grupului de control. Următoarele trei persoane care au intrat în studiu au toate numere pare și sunt în

grupul experimental. Pe termen lung, va exista un număr egal de pacienți în fiecare dintre cele două grupuri.

6.4 Teste cu două cozi versus test cu o singură coadă

Un studiu clinic este conceput pentru a testa o anumită ipoteză. Se vede adesea această frază în articolele de cercetare: „Semnificativ la nivel .05, test cu două cozi". Amintiți-vă că într-o secțiune anterioară, am discutat conceptul de „ipoteză nulă", care afirmă că nu există nicio diferență între două grupuri pe o măsură de interes. Am spus că, pentru a testa această ipoteză , vom aduna date astfel încât să putem decide dacă ar trebui să respingem ipoteza lipsei de diferență în favoarea unei ipoteze alternative . Un test cu două cozi versus un test cu o singură coadă se referă la ipoteza alternativă propusă. De exemplu, să presupunem că sunteți interesat să comparați nivelul mediu de colesterol al unui grup tratat cu un medicament pentru scăderea colesterolului cu media unui grup de control căruia i sa administrat un placebo. Veți colecta datele adecvate dintr-un studiu bine conceput și ați configura ipoteza nulă ca

H_o : Colesterol mediu în grupul tratat = colesterol mediu în con grupul de control

Puteți alege ca ipoteză alternativă

H_A : Colesterolul mediu în grupul tratat este mai mare decât media în controale

În această împrejurare, ați respinge ipoteza nulă în favoarea ipotezei alternative dacă media observată pentru femei ar fi suficient mai mare decât media observată pentru bărbați, pentru a vă conduce la

 concluzia că o diferență atât de mare în această direcție nu este probabil să se fi produs doar întâmplător. Acesta, deci, ar fi un test cu o singură coadă al ipotezei nule.

Dacă, totuși, ipoteza dvs. alternativă a fost că nivelul mediu de colesterol pentru femei este diferit de nivelul mediu de colesterol pentru bărbați, atunci ați respinge ipoteza nulă în favoarea alternativei fie dacă media pentru femei a fost suficient de mai mare decât media pentru bărbați sau dacă media pentru femei a fost suficient de mai mică decât media pentru bărbați. Direcția diferenței nu este specificată. În cercetarea medicală, este mai obișnuit să folosim un test de semnificație cu două cozi, deoarece adesea nu știm în ce direcție poate fi o diferență, chiar dacă putem crede că știm înainte de a începe experimentul. În orice caz, este important să raportăm dacă folosim un test cu o coadă sau un test cu două cozi.

6.5 Studiul clinic ca „ Standard de aur ”

Uneori, dovezile din studiile observaționale pot duce la concluzii înșelătoare cu privire la eficacitatea sau siguranța unui tratament, doar pentru a fi anulate de dovezile din studiile clinice, cu implicații enorme pentru sănătatea publică. Studiul clinic al terapiei hormonale a Inițiativei pentru sănătatea femeilor (WHI) este un exemplu dramatic în acest sens. [22] Estrogenul a fost aprobat de FDA pentru ameliorarea simptomelor postmenopauzei în 1942, comercializat în mod agresiv la mijlocul anilor 1960, iar după 1980, în general, combinat cu progestativ pentru femeile cu uter, deoarece s-a descoperit că progestinul compensa riscurile de estrogen pentru cancerul uterin. . Între timp, multe studii prospective mari de urmărire au arătat aproape uniform că estrogenul a redus bolile de inimă cu 30-50 %. În 1993, WHI a organizat un studiu clinic amplu pentru a răspunde cu adevărat la întrebarea privind riscurile și beneficiile pe termen lung ale terapiei hormonale. O parte a fost studiul numai estrogenului pentru femeile care au suferit o histerectomie și, prin urmare, nu au avut nevoie de progesteron pentru a-și proteja uterul, iar o altă parte a fost de estrogen plus progesteron (E + P) pentru femeile cu uterul intact.

Studiul E + P a fost un studiu clinic randomizat, dublu-orb, controlat cu placebo, menit să se desfășoare pentru o medie de 8,5 ani. Acesta a inclus 16.608

femei între 50-79 de ani; o dimensiune atât de mare a eșantionului a fost considerată necesară pentru a obține o putere adecvată. Procesul a fost oprit în 2002, cu 3 ani înainte de finalizarea sa planificată, deoarece Consiliul de monitorizare a datelor și siguranței sau DSMB (așa cum este descris în capitolul 10) au descoperit că estrogenul plus progestativ au cauzat un exces de cancer de sân și, în mod surprinzător, a existat un exces semnificativ și complet neașteptat de atacuri de cord în grupul E + P în comparație cu placebo! Rezultatele finale, raportate în lucrările ulterioare, au arătat că efectele adverse (o creștere cu 24% a cancerului de sân invaziv, creșterea cu 31% a accidentelor vasculare cerebrale, creșterea cu 29% a bolilor coronariene și o creștere de peste două ori a emboliei pulmonare și a demenței) compensarea beneficiilor (o scădere cu 37 % a cancerului colorectal și scădere cu 34 % a fracturilor de șold), astfel încât, luate împreună, numărul de evenimente dăunătoare în exces pe an a fost substanțială. Deoarece au existat șase milioane de femei care au luat acest preparat numai în Statele Unite și alte milioane la nivel global, aceste rezultate au implicații importante pentru alte femei decât cele din studiu în sine.

De ce rezultate atât de diferite dintr-un studiu clinic decât din studii longitudinale observaționale? Cea mai probabilă explicație este prejudecata de selecție. Femeile care luau hormoni și apoi urmau să-și observe ratele bolilor de inimă erau, practic în toate studiile observaționale, mai sănătoase, mai slabe, mai active, mai educate și mai puțin supraponderale decât omologii lor care nu iau hormoni și mai sănătoase. stilul de viață și starea de sănătate de bază mai bună, mai degrabă decât hormonii în sine, a fost ceea ce a explicat ratele lor mai scăzute de boli de inimă.

La întrebarea se răspunde acum folosind „standardul de aur”, studiul clinic: estrogenul plus progestativ nu protejează împotriva bolilor de inimă și de fapt crește riscul. După cum sa menționat anterior, impactul acestei cercetări este mare, deoarece atât de multe milioane de femei foloseau preparatul testat.

6.6 Regresia spre medie

Când selectați dintr-o populație acei indivizi care au hipertensiune arterială și apoi, ulterior, măsurați din nou tensiunea arterială, media celor de-a doua măsurători va tinde să fie mai mică decât

media primelor măsurători și va fi mai aproape de media populației inițiale din care au fost extrași acești indivizi. Dacă între prima și a doua măsurători ați instituit un tratament, este posibil să atribuiți incorect scăderea tensiunii arteriale medii în grup efectelor tratamentului, în timp ce o parte a acestei scăderi se poate datora fenomenului numit regresie spre medie. (Acesta este unul dintre motivele pentru care un grup de control placebo este cel mai important pentru compararea efectelor tratamentului peste și dincolo de cele cauzate de regresia la medie.) Regresia la medie apare atunci când selectați un grup deoarece indivizii au valori care se situează peste unele nivel de criteriu, ca în screening. Se datorează variabilității erorii de măsurare. Luați în considerare tensiunea arterială.

Valoarea observată a tensiunii arteriale este valoarea adevărată a persoanei plus o cantitate necunoscută de eroare. Presupunerea este că tensiunea arterială măsurată de oameni este în mod normal distribuită în jurul valorii medii a valorii lor adevărate, dar necunoscute, a tensiunii arteriale. Să presupunem că vom include oameni în studiul nostru doar dacă tensiunea arterială este de 160 sau mai mult. Acum să presupunem că tensiunea arterială sistolica adevărată a cuiva este 150, dar o măsurăm 160. Selectăm acea persoană pentru grupul nostru de studiu doar pentru că valoarea sa măsurată este mare. Cu toate acestea, data viitoare când îi vom măsura tensiunea arterială, probabil că el va fi mai aproape de valoarea sa reală de 150 decât prima dată. (Dacă ar fi fost aproape de valoarea sa adevărată de 150 prima dată, nu l-am fi selectat niciodată pentru studiul nostru, pentru că ar fi fost sub punctul nostru de limită. Așa că trebuie să fi avut o eroare mare la început. Deoarece aceste erori sunt distribuite în mod normal în jurul valorii medii reale de 150, data viitoare avem mai multe șanse să obținem o eroare mai mică și, prin urmare, o tensiune arterială măsurată mai mică decât 160 care ne-au determinat să-l selectăm.

Să presupunem acum că selectăm un întreg subgrup de oameni care au valori ridicate. Mediile celor doua măsurători ale acestor persoane selectate vor tinde să fie mai mici decât media primelor lor măsurători și mai apropiate de media întregului grup din care i-am selectat. Ideea este că persoanele care au cele mai mari valori prima dată nu au întotdeauna cele mai mari valori a doua oară deoarece corelația dintre prima și a doua măsurătoare nu este perfectă. În mod similar, dacă selectăm un grup de persoane din cauza valorilor scăzute ale unei anumite caracteristici, media celei de-a doua măsurători la aceste persoane va fi mai mare. decât media primelor lor măsurători și din nou mai aproape de media întregului grup.

O altă explicație a acestui fenomen poate fi ilustrată prin următorul exemplu de aruncare a unui zar. Imaginează-ți că arunci un zar de 360 de ori. Ori de câte ori zarul ajunge pe un cinci sau un șase, vei arunca din nou zarul. Suntem interesați de trei medii diferite: (1) media primelor 360 de aruncări, (2) media aruncărilor care vor avea ca rezultat aruncarea noastră din nou și (3) media celor de-a doua aruncări. Rezultatele noastre sunt prezentate în tabelul de mai jos.

Deși la prima aruncare, media celor 360 de aruncări este 3,5, alegem doar cele două numere cele mai mari, iar media lor este 5,5. Aceste 120 de ori când zarul a aterizat pe 5 sau 6 ne vor face să aruncăm din nou, dar la a doua aruncare, rezultatul poate varia în mod liber între 1 și 6. Prin urmare, media celei de-a doua aruncări trebuie să fie mai mică decât media a doua aruncare. grup pe care l-am selectat din prima aruncare în mod special pentru că avea valorile mari.

<table>
<tr><td colspan="2">Prima aruncare</td><td colspan="2">A doua aruncare</td></tr>
<tr><td>Rezultat</td><td># de ori se obține rezultatul</td><td>Rezultat</td><td># de ori se obține rezultatul</td></tr>
<tr><td>1</td><td>60</td><td></td><td></td></tr>
<tr><td>2</td><td>60</td><td></td><td></td></tr>
<tr><td>3</td><td>60</td><td></td><td></td></tr>
<tr><td>4</td><td>60</td><td></td><td></td></tr>
<tr><td>5</td><td>60</td><td>1</td><td>20</td></tr>
<tr><td>6</td><td>60</td><td>2</td><td>20</td></tr>
<tr><td></td><td></td><td>3</td><td>20</td></tr>
<tr><td></td><td></td><td>4</td><td>20</td></tr>
<tr><td></td><td></td><td>5</td><td>20</td></tr>
<tr><td colspan="2">Media a 360 de aruncări = 3,5</td><td>6</td><td>20</td></tr>
<tr><td colspan="2">Media celor 120 de aruncări care au aterizat 5 sau 6 = 5,5</td><td colspan="2">Media a 2-a aruncare = 3,5</td></tr>
</table>

6.7 Analiza intentiei de tratare

Datele din studiile clinice în general ar trebui analizate prin compararea grupurilor așa cum au fost inițial randomizate și nu prin compararea cu grupul de control placebo doar a celor din grupul de droguri care au luat efectiv medicamentul. Persoanele alocate grupului de droguri active ar trebui incluse în acel grup pentru

analiză, chiar dacă nu au luat niciodată medicamentul. Acest lucru poate suna ciudat, deoarece cum se poate evalua eficacitatea unui medicament dacă pacientul nu îl ia? Dar tocmai motivul pentru care oamenii ar putea să nu se conformeze regimului de medicamente poate avea de-a face cu efectele adverse ale medicamentului, astfel încât, dacă îi selectăm numai pe cei care se conformează, avem un grup diferit de cel randomizat și putem avea un imagine părtinitoare a efectelor drogurilor.

Un alt aspect este că poate exista o anumită calitate a complianților în general care afectează rezultatul. Un exemplu faimos de concluzii înșelătoare care ar putea apărea din neefectuarea unei analize intenționate de tratare provine din Proiectul de medicamente coronariene. [28] Acest studiu randomizat, dublu-orb, a comparat medicamentul clofibrat cu placebo pentru reducerea colesterolului. Variabila rezultatului, care a fost mortalitatea la 5 ani, a fost foarte similară în ambele grupuri, 18% în grupul cu medicamente și 19% în grupul placebo. S-a dovedit, totuși, că doar aproximativ două treimi dintre pacienții care ar fi trebuit să ia clofibrat au fost de fapt complianți și și-au luat medicamentele . Acești oameni au avut o rată a mortalității de 15%, semnificativ mai mică decât mortalitatea de 19% în grupul placebo. Cu toate acestea, o analiză ulterioară a arătat că dintre cei alocați grupului placebo, o treime nu și -a luat nici pastilele placebo. Cele două treimi din grupul placebo care s-au conformat au avut o mortalitate de 15%, la fel ca și cei care au respectat medicamentul clofibrat! Persoanele neconforme din grupul de droguri și placebo au avut o mortalitate mai mare (25% pentru clofibrat și 28% pentru placebo). Poate fi de dorit în anumite circumstanțe să se analizeze efectul unui medicament la cei care îl iau efectiv. În acest caz, comparația dintre cei care respectă medicamentele ar trebui să fie mai degrabă cu cei care respectă placebo decât cu grupul placebo în ansamblu.

Includerea persoanelor care nu respectă în analiză diluează efectele, așa că ar trebui depuse toate eforturile pentru a minimiza neconformitatea. În unele studii, capacitatea evaluată de conformare este un criteriu de înscriere și se face o evaluare a fiecărui pacient ca parte a stabilirii acestuia.
eligibilitatea cu privire la probabilitatea ca acest pacient să adere la regimul. Cei care nu ar putea face acest lucru sunt excluși înainte de randomizare. Cu toate acestea, dacă întrebarea la îndemână este cât de acceptabil este tratamentul pentru pacient, pe lângă eficacitatea acestuia, atunci baza pentru includere poate fi populația generală care ar putea beneficia de medicament, inclusiv cei care nu respectă.

În Inițiativa pentru sănătatea femeilor , analiza principală a fost intenția de a trata. Cu toate acestea, o analiză secundară ajustată pentru conformitate (denumită mai frecvent aderență). În această analiză, istoricul evenimentului al participantei a fost cenzurat la 6 luni după ce ea fie a încetat să mai ia pastilele de studiu, fie a luat mai puțin de 80% din pastilele de studiu. În grupul placebo, istoricul evenimentului a fost cenzurat la 6 luni după ce participantul a început să ia hormoni (unii participanți din grupul placebo au încetat să ia pastilele de studiu, dar li s-au prescris hormoni de către medici și au început să le ia singuri). Astfel, această analiză secundară a comparat în principiu cele două grupuri „așa cum au fost tratate", mai degrabă decât așa cum au fost atribuite unui anumit tratament. În analiza intenției de tratament, raportul de risc pentru boala coronariană a fost de 1,24, în timp ce în analiza „ajustată în funcție de aderență", acesta a fost de 1,50. Astfel, constatările din analiza intention-to-treat au fost confirmate și consolidate în analizele ajustate în funcție de aderență.

6.8 Cât de mare ar trebui să fie studiul clinic?

Un studiu clinic ar trebui să fie suficient de mare, adică să aibă o dimensiune a eșantionului suficient de mare, pentru a avea o probabilitate mare de a detecta o diferență reală între cele două grupuri. Dacă faceți o mică încercare și nu găsiți nicio diferență semnificativă, nu ați obținut informații noi; este posibil să nu fi găsit o diferență pur și simplu pentru că nu ai avut destui oameni în studiu. Nu puteți afirma că nu există nicio diferență între tratamente. Dacă aveți un proces mare și nu găsiți nicio diferență semnificativă, atunci puteți spune cu mai multă certitudine că tratamentele nu sunt într-adevăr diferite.

Să presupunem că găsiți o diferență semnificativă într-un studiu mic cu $p < .05$ (nivel de semnificație). Aceasta înseamnă că rezultatul pe care l-ați obținut este probabil să fi apărut pur întâmplător de mai puțin de 5 ori din 100 (dacă nu ar exista într-adevăr nicio diferență). Este de încredere la fel de mult ca aceeași valoare p dintr-un proces mare? Există mai multe școli de gândire despre asta.

Valoarea p este un indice al puterii dovezilor cu privire la respingerea unei ipoteze nule. Unii oameni cred că valoarea ap este valoarea ap și are aceeași greutate, indiferent dacă provine dintr-un studiu mare sau mic. Alții cred că, dacă obțineți un rezultat semnificativ într-un studiu mic, înseamnă că efectul (sau

diferența dintre două medii ale populației) trebuie să fie suficient de mare încât să puteți detecta chiar și cu eșantioanele dvs. mici și, prin urmare, este este o diferență semnificativă. Este adevărat că, dacă dimensiunea eșantionului este suficient de mare, putem găsi semnificație statistică dacă diferența reală dintre medii este foarte, foarte mică și practic irelevantă. Prin urmare, găsirea unei diferențe semnificative într-un studiu mic înseamnă că efectul a fost relativ mare.

Alții spun că, în practică, totuși, puteți avea mai puțină încredere că tratamentele diferă într-adevăr pentru o anumită valoare p într-un studiu mic decât dacă ați fi obținut aceeași valoare p la testarea acestor două tratamente într-un studiu mare. [29] Acest aparent paradox poate apărea în situațiile în care există mult mai multe studii mici care se desfășoară în întreaga lume care studiază aceeași problemă decât există studii mari, cum ar fi în terapia cancerului. Unele dintre aceste încercări, doar întâmplător, se vor dovedi a avea rezultate semnificative care pot induce în eroare.

Să presupunem că există 1.000 de teste mici de terapie cu medicamente anticancer. Numai întâmplător, 5% dintre acestea vor fi semnificative chiar dacă terapiile nu au niciun efect, sau 50 de rezultate semnificative. Deoarece acestea sunt doar întâmplătoare, înseamnă că suntem incorecți să declarăm efectele medicamentelor anticanceroase în aceste studii (am comis erori de tip I). Să presupunem, în plus, că există doar 100 de studii mari care studiază aceeași problemă. Dintre acestea, 5 %, sau cinci astfel de studii, vor declara că există o diferență, în mod incorect. Deci, dacă combinăm incorect toate încercările care arată diferențe semnificative, avem 55 de astfel de valori p semnificative, dar înșelătoare. Dintre acestea, 50% sau 91% provin din încercări mici, iar 5 din cele 55 incorecte (sau 9%) provin din încercări mari. Următoarele puncte sunt importante:

(1) Există o distincție între semnificația statistică și semnificația clinică. Un rezultat poate să nu fi apărut întâmplător, adică să reflecte o diferență adevărată, dar să fie atât de mic încât să nu-l facă să nu aibă importanță practică.

(2) Cel mai bine este să raportați probabilitatea reală de a obține rezultatul doar întâmplător sub ipoteza nulă, adică valoarea reală p, mai degrabă decât să spuneți că este semnificativ sau nu. Valoarea p pentru ceea ce numim în mod obișnuit „semnificație" este arbitrară. Prin obicei, sa considerat a fi o valoare ap de .05 sau mai puțin. Dar punctul de limită .05 nu este sacru. Cititorul ar trebui să decidă ce putere va pune în

dovezile furnizate de studiu, iar cititorul trebuie să aibă informațiile necesare pentru a lua acea decizie. Informațiile trebuie să includă proiectarea studiului, selecția eșantionului, dimensiunile eșantionului, abaterile standard și valorile reale ale p.

În concluzie :

(1) Găsirea unei diferențe semnificative dintr-un mic proces nu ne spune nimic.
(2) Găsirea unei diferențe semnificative într-un studiu mare este o constatare reală și ne spune că este posibil ca tratamentele să fie echivalente.
(3) Găsirea unei diferențe semnificative într-un studiu mic poate fi sau nu replicabilă.
(4) Găsirea unei diferențe semnificative într-un proces mare este de încredere ca dezvăluirea unei adevărate diferențe.

6.9 Ce este implicat în calculul dimensiunii eșantionului?

(a) Dimensiunea efectului

Să spunem că 15 % dintre victimele unui anumit tip de atac de cord mor dacă li se administrează medicamentul A și 16 % mor dacă li se administrează medicamentul B. Această diferență de 1 % înseamnă că medicamentul A este mai bun? Majoritatea oamenilor ar spune că aceasta este o diferență prea mică, chiar dacă nu apare întâmplător, pentru a avea vreo importanță clinică. Să presupunem că diferența dintre cele două medicamente este de 5%. Am spune acum că medicamentul A este mai bun? Asta ar depinde de cât de mare a

 diferența pe care am considerat-o importantă. Mărimea diferenței pe care vrem să o detectăm se numește dimensiunea efectului.

Pentru a calcula dimensiunea eșantionului, trebuie să cunoașteți dimensiunea minimă a diferenței dintre două tratamente pe care ați fi dispus să nu le detectați. Să presupunem, de exemplu, că în grupul dumneavoastră de control, 30 % dintre pacienții fără tratament se recuperează. Convingerea dumneavoastră este că cu tratamentul în grupul experimental, 40% se vor recupera. Crezi că această diferență în rata de recuperare este importantă din punct de vedere clinic și vrei să fii sigur că poți detecta o diferență cel puțin la fel de mare ca diferența dintre 30 % și 40 %. Aceasta înseamnă că, dacă rata de recuperare a grupului de tratament ar fi de 35%, ați fi dispus să nu găsiți un efect atât de mic. Cu toate acestea, dacă rata de tratament a fost de 40% sau mai mult, ați dori să fiți destul de sigur că o găsiți. Cât de sigur ai vrea să fii? Problema „cât de sigur" are de-a face cu „puterea" testului statistic.

(b) Putere

Puterea statistică înseamnă probabilitatea de a găsi un efect real (de mărimea pe care o considerați importantă clinic). Relațiile dintre putere, nivelul de semnificație și eroarea de tip I și tip II sunt rezumate mai jos:

Nivel de semnificație = probabilitatea unei erori de tip I = probabilitatea de a găsi un efect atunci când într-adevăr nu există . Acest lucru este cunoscut și sub numele de alfa sau α .

Probabilitatea erorii de tip II = probabilitatea de a nu găsi un efect atunci când există într-adevăr unul. Acest lucru este cunoscut și sub denumirea de beta sau β .

Puterea = 1-probabilitate de eroare de tip II = probabilitatea de a găsi un efect atunci când există într-adevăr unul. Acesta este, de asemenea, cunoscut sub numele de 1 – *beta.*

(c) Dimensiunea eșantionului

Pentru a calcula dimensiunea eșantionului, trebuie să specificați dimensiunea efectului, nivelul de semnificație și puterea dorită . Dacă alegeți un nivel de semnificație de .05 și o putere de .80, atunci probabilitatea dvs. de eroare de tip II este 1 - putere sau .20. Aceasta înseamnă că considerați o eroare de tip I ca fiind

de patru ori mai gravă decât o eroare de tip II (.20/.05 = 4) sau că vă este de patru ori mai frică să găsiți ceva care nu există decât să nu găsiți. ceva care este.

Când calculați dimensiunea eșantionului, există întotdeauna un compromis. Dacă doriți să reduceți probabilitatea de a face o eroare de tip I, atunci pentru o anumită dimensiune a eșantionului și dimensiunea efectului, veți crește probabilitatea de a face o eroare de tip II. Puteți menține la un nivel scăzut ambele tipuri de erori prin creșterea dimensiunii eșantionului. Partea de sus a tabelului de pe pagina următoare arată dimensiunile eșantionului necesare pentru a compara două grupuri cu un test între două proporții în ipoteze diferite.

Al doilea rând al tabelului arată că, dacă doriți să puteți detecta o diferență în rata de răspuns de la 30 % în grupul de control la 50 % sau mai mult în grupul de tratament cu o probabilitate (putere) de 0,80, ar trebui 73 de persoane în fiecare dintre cele două grupuri. Dacă, totuși, vrei să fii destul de sigur că găsești o diferență la fel de mică precum cea dintre 30 % și 40 %, atunci trebuie să ai 280 de persoane în fiecare grup.

Dacă doriți să fiți mai sigur că găsiți diferența, să spunem 90% sigur în loc de 80% sigur, atunci veți avea nevoie de 388 de persoane în fiecare grup (mai degrabă decât 280 pentru putere .80). Dacă doriți să aveți un nivel de semnificație mai strict de .01, veți avea nevoie de 118 persoane în fiecare grup (comparativ cu cele 73 necesare pentru nivelul de semnificație .05) pentru a putea detecta diferența dintre 30 % și 50 %; veți avea nevoie de 455 de persoane (comparativ cu 280 pentru nivelul .05) pentru a detecta o diferență de la 30 % la 40 % rata de răspuns.

Partea de jos a tabelului de pe pagina următoare arată impactul asupra dimensiunii eșantionului al unui test de semnificație cu o singură coadă față de un test cu două cozi. Amintiți-vă că un test cu două cozi postulează că rata de răspuns în grupul de tratament poate fi fie mai mare, fie mai mică decât rata de răspuns în grupul de control, în timp ce un test cu o singură coadă specifică direcția diferenței ipotetice. Un test cu două cozi necesită o dimensiune mai mare a eșantionului, dar acesta este cel mai frecvent utilizat.

(d) Câteva considerații suplimentare

Pentru o dimensiune fixă a eșantionului și o anumită dimensiune a efectului sau diferență pe care doriți să o detectați, puterea maximă apare atunci când rata evenimentului este de aproximativ 50 %. Deci, pentru a maximiza puterea, poate fi uneori înțelept să selectați un grup pentru studiu care este probabil să aibă evenimentele de interes. De exemplu, dacă doriți să studiați efectele unui medicament beta-blocant asupra prevenirii atacurilor de cord, puteți obține „mai multă putere pentru bani” studiind persoanele care au deja

a avut un atac de cord mai degrabă decât persoanele sănătoase, deoarece primele sunt mai susceptibile de a avea un alt eveniment (atac de cord). Desigur, atunci s-ar putea să vă uitați la o altă întrebare, efectul beta-blocantelor asupra supraviețuitorilor atacului de cord (care ar fi o cale de prevenire secundară), mai degrabă decât efectul beta-blocantelor în prevenirea primului atac de cord (un studiu de prevenire primară). Uneori, un studiu de prevenire primară oferă un răspuns diferit decât un studiu de prevenire secundară. Este posibil să puteți interveni pentru a preveni îmbolnăvirea în rândul persoanelor care nu suferă încă de boală, dar intervenția dvs. poate avea un efect redus asupra unei persoane care a dezvoltat deja boala. În mod clar, este nevoie de judecată.

Exemple de dimensiuni ale mostrelor

Nivel de semnificație (1 coadă)	Să presupunem	Dimensiunea efectului	Putere	Dimensiunea eșantionului
	Controla grup răspuns rata =	Detectează creșterea în grup de tratament la cel puțin sa	Cu probabilitate de	n Necesar în fiecare grup
.05	30 % 30 %	40 % 50 %	.80 .80	280 73
	30 % 30 %	40 % 50 %	.90 .90	388 101
.01	30 % 30 %	40 % 50 %	.80 .80	455 118
	30 % 30 %	40 % 40 %	.90 .90	590 153

Exemple de dimensiuni ale mostrelor

Nivel de semnificație = .05	Să presupunem	Dimensiunea efectului	Putere	Dimensiunea
	Rata de răspuns a grupului de control =	Detectează creșterea grupului de tratament cel puțin până la	Cu probabilitate de	*n* Necesar în fiecare grup
1-coada 2-cozi	30 % 30 %	40 % 40 %	.80 .80	280 356

1-coada 2-cozi	30 % 30 %	50 % 50 %	.80 .80	73 92

6.10 Cum se calculează dimensiunea eșantionului pentru diferența dintre două proporții

Trebuie să specificați care credeți că este probabil să fie proporția de evenimente în fiecare dintre cele două grupuri comparate. Un eveniment poate fi un răspuns, o moarte sau o recuperare, dar trebuie să fie o variabilă dihotomică. Specificațiile dvs. privind ratele evenimentelor din cele două grupuri reflectă dimensiunea diferenței pe care ați dori să o puteți detecta.

Specifica:

p_1 = rata in grupa 1 ; $q_1 = 1 - p_1$; alfa = nivel de semnificație

p_2 = rata in grupa 2 ; $q_2 = 1 - p_2$; putere

$n = (pq_1) + (i\frac{1}{8}{}^{q}2) \times f$ *ðaipha , putere)*

&2 $-P_1)$

Valorile lui *f* (alfa, putere) pentru un test cu două cozi pot fi obținute din tabelul de mai jos.

Values of *f* (alpha, power)

		.95	.90	.80	.50
Alpha significance level	.10	10.8	8.6	6.2	2.7
	.05	13.0	10.5	7.9	3.8
	.01	17.8	14.9	11.7	6.6

Notă: n este aproximativ invers proporțional cu $(p_2 - p_1)^2$.

Exemplu Să presupunem că doriți să găsiți dimensiunea eșantionului pentru a detecta o diferență de la 30 % la 40 % între două grupuri, cu o putere de 0,80 și un nivel de semnificație de 0,05. Apoi,

$$p_1 = : 30 \; ; \; q_1 = : 70 \; ; \; \text{alfa} = : 05$$

$$p_2 = : 40 \; ; \; q_2 = : 60 \; ; \; \text{putere} = : 80$$

$$f(\text{alfa, putere}) = 7 . 9 \text{ de la masă}$$

$$\frac{(. 30)(. 70) + (. 40)(. 60)}{(: 40 - : 30)^2} \quad X \; 7{:}9 = 356$$

Veți avea nevoie de 356 de persoane în fiecare grup pentru a fi 80% sigur că puteți detecta o diferență de la 30% la 40% la nivelul .05.

6.11 Cum se calculează dimensiunea eșantionului pentru a testa diferența dintre două mijloace

Formula pentru calcularea mărimii eșantionului pentru un test al diferenței dintre două medii, presupunând că va exista un număr egal în fiecare grup, este

$$n = \frac{k^2 \sigma_2}{(MD)^2} = \text{număr în fiecare grupă}$$

unde σ^2 este varianța erorii, *MD* este diferența minimă dorită

detectează, iar k depinde de nivelul de semnificație și puterea dorită. Valorile selectate ale lui k sunt afișate pe pagina următoare. De exemplu, pentru a detecta o diferență în IQ-ul mediu de 5 puncte între două grupuri de oameni, unde varianța = $16^2 = 256$, la un nivel de semnificație de .05 și cu puterea de .80, am avea nevoie de 161 în fiecare grup, sau o dimensiune totală a eșantionului de 322. Aceasta înseamnă că avem 80 % probabilitatea de a detecta o diferență mai mare sau mai mare de 5 puncte. Pentru o diferență de 10 puncte, am avea nevoie de 54 de persoane în fiecare grup.

$$n = \frac{7 : 849 \times 2\,(256)}{(5)^2} = 161 \text{ persoane}$$

Nivel de semnificație	Putere	k
.05	.99	18.372
	.95	12.995
	.90	10.507
	.80	7.849
.01	.99	24.031
	.95	17.814
	.90	14.879
	.80	11.679

Un set comun de parametri pentru astfel de calcule ale dimensiunii eșantionului sunt $\alpha = .05$ și puterea = .80. Cu toate acestea, atunci când există mai multe comparații, trebuie să setăm α la niveluri inferioare, așa cum este descris în secțiune 3.24 asupra procedurii Bonferroni. Apoi, dimensiunea eșantionului nostru ar trebui să fie mai mare.

Dacă sperăm să arătăm că două tratamente sunt echivalente, trebuie să setăm diferența minimă pe care dorim să o detectăm să fie foarte mică și puterea să fie foarte, foarte mare, rezultând dimensiuni foarte mari ale eșantionului.

Pentru a calcula valorile lui k care nu sunt tabulate aici, cititorul este trimis la cartea Methods in Observational Epidemiology de Kelsey, Whittmore, Evans și Thompson pentru o explicație excelentă. Există programe de calculator disponibile pentru a calcula puterea pentru multe situații diferite . Un excelent este NCSS (software statistic al Consiliului Național pentru Studii Sociale) care poate fi

obținut accesând site-ul www. ncss.com .

Chapter 7
CELE MAI MULT DESPRE CALITATEA VIEI

Viața care este neexaminată nu merită trăită.

Platon, Dialoguri
428–348 î.Hr

Îmi place viața lungă mai mult decât smochinele.

Shakespeare
(Anthony și Cleopatra)

Cele două citate de mai sus ilustrează cât de diferit văd oamenii calitatea vieții lor și cât de dificil este să stabilești acest concept.

O dezvoltare binevenită în cercetarea în domeniul sănătății este atenția tot mai mare acordată problemelor de calitate a vieții în studiile epidemiologice și atunci când se evaluează terapiile concurente. Un aspect cheie este măsurarea efectelor simptomelor de boală, precum și a tratamentului acestor simptome, asupra bunăstării, care este o stare subiectivă și relativă. Prin urmare, este destul de adecvat ca măsurarea îmbunătățirii sau a deteriorării calității vieții să se bazeze pe percepția și auto-raportarea pacientului . O persoană care a avut angină severă și invalidantă poate percepe o stare de bine îmbunătățită ca urmare a tratamentului dacă poate merge fără durere, în timp ce un tânăr pasionat de schi poate experimenta o deteriorare marcată dacă nu poate schia. Din acest motiv, în studiile pe această problemă, individul servește adesea drept propriul control, iar măsurile utilizate sunt scorurile de modificare în anumite dimensiuni ale calității vieții de la înainte până la după tratament. Cu toate acestea, rămâne important să existe un grup de control adecvat pentru a compara modificările, deoarece oamenii prezintă modificări ale acestor dimensiuni în timp, care pot fi fără legătură cu tratamentul particular evaluat.

Principiile și tehnicile descrise în această carte se aplică cercetării în orice domeniu legat de sănătate. Cu toate acestea, există anumite analitice

metode care sunt deosebit de adecvate investigațiilor privind stările psihologice

sau emoționale. Principiul de bază este că, dacă urmează să fie cercetare științifică, aceasta trebuie să respecte standardele științifice, ceea ce înseamnă că în primul rând variabilele de interes trebuie cuantificate. Din fericire, aproape orice concept legat de domeniile sănătății poate fi cuantificat dacă cineva este suficient de ingenios.

7.1 Construcție la scară

Scalele utilizate pentru măsurarea dimensiunilor calității vieții reflectă gradul de suferință cu anumite simptome sau stări psihologice, precum și gradul de satisfacție și bunăstare generală. Există multe astfel de scale disponibile , care au fost bine construite și testate pe diferite populații. Uneori, totuși, anchetatorii consideră că este necesar să-și construiască propriile scale pentru a se potrivi circumstanțelor particulare.

Există trei caracteristici ale unor astfel de scale care sunt importante: fiabilitatea, validitatea și capacitatea de răspuns.

7.2 Fiabilitate

Fiabilitatea este capacitatea de a măsura ceva în același mod de două ori. Se bazează pe ipoteza că scorul unei persoane pe o scară sau un test este compus din scorul său adevărat (dar necunoscut) plus o componentă care este supusă variației din cauza erorii (prin care înțelegem variabilitatea aleatorie).

Fiabilitatea unei scale este legată de repetabilitatea acesteia sau de cât de apropiate sunt răspunsurile la două administrări ale scalei. Pentru a măsura cât de apropiate sunt, putem calcula coeficientul de corelație dintre cele două administrări ale scalei la aceiași subiecți. Dar adesea nu putem da aceeași scară pacienților noștri de două ori exact în aceleași circumstanțe, deoarece, în realitate, un pacient care răspunde de două ori la aceleași întrebări ar răspunde diferit fie pentru că a intervenit ceva între cele două ocazii, fie pentru că și-a amintit răspunsurile anterioare. sau doar pentru că există o variabilitate inerentă în modul în care te simți. Următorul cel mai bun lucru

ar fi să dai două scale echivalente aceluiași grup, dar și asta are problemele sale. De unde știm că cele două scale sunt cu adevărat echivalente?

Din fericire, există diverse măsuri ale ceea ce numim „coerență internă" care ne oferă fiabilitatea unei scale sau a unui test. Cel mai comun se numește alfa lui Cronbach . Există multe pachete de software pentru computerele personale care calculează cu ușurință alfa lui Cronbach , inclusiv SPSS, SAS, STATA și multe altele. Astfel, nu este necesar să îl calculați singur, dar următoarea explicație indică ce înseamnă cu adevărat și cum să o interpretați.

$$\frac{k}{k-1}\left(\frac{\text{variația scării totale} - \text{suma variațelor elementelor individuale}}{\text{Varianceoftotalscale}}\right)$$

Varianta este abaterea standard la pătrat. Secțiune 3.4 arată cum se calculează. (Când vorbim despre varianță aici, ne referim de fapt la varianța populației, dar ceea ce folosim cu adevărat sunt estimări ale variației populației pe care le obținem din eșantionul special de oameni pe care dezvoltăm testul sau scara, deoarece, evident, putem " nu măsoară întreaga populație.)

Această formulă este într-adevăr o măsură a cât de omogene sunt elementele de scară, adică în ce măsură măsoară același lucru. Dacă aveți o scară care este compusă din mai multe subscale diferite, fiecare măsurând lucruri diferite, atunci alfa lui Cronbach ar trebui utilizată pentru fiecare dintre subscale separat, mai degrabă decât întreaga scară. Alfa-ul lui Cronbach oferă limita inferioară pentru fiabilitate. Dacă este mare pentru întreaga scală, atunci știți că scala este de încredere (repetabilă, foarte corelată cu scorurile „adevărate", dar necunoscute). Dacă obțineți un alfa scăzut pentru întreaga scară, atunci fie nu este de încredere, fie măsoară mai multe lucruri diferite.

Fiabilitatea poate fi privită și ca o măsură a corelației și, de fapt, reflectă corelația medie între itemii unei scale, luând în considerare numărul total de itemi. O altă modalitate de a obține fiabilitate este din formula Spearman-Brown, care este

$$\frac{k\,(\text{corelația medie între toate articolele})}{1 + (k-1)\,\text{corelația medie între toate elementele}} \quad \frac{k(r_{medie})}{1 + (k-1)(r_{media})}$$

După cum indică această formulă, un test sau o scară mai lungă este în general mai de încredere dacă elementele suplimentare măsoară același lucru. Pe de altă parte,

scalele mai scurte sunt mai acceptabile pentru pacienți. Un alfa peste .80 este considerat foarte bun, iar uneori subscalele sunt acceptabile cu alfa peste .50, mai ales atunci când există un număr mare de subiecți (peste 300), dar ar trebui luat în considerare în contextul celorlalte calități psihometrice ale scara.

Există și alte măsuri de fiabilitate. Psihometria este un domeniu specializat și complex și există multe cărți excelente pe acest subiect, de exemplu, *Scale de măsurare a sănătății* de Streiner și Norman.

7.3 Valabilitate

Valabilitatea se referă la gradul în care testul măsoară ceea ce ar trebui să măsoare. O situație ideală ar fi aceea în care există un criteriu extern în funcție de care să judeci instrumentul de măsurare, un „etalon de aur". De exemplu, dacă s-ar putea demonstra că anxietatea măsurată pe o scară se corelează mai bine cu un rezultat definibil și convenit în mod obiectiv decât anxietatea măsurată pe o a doua scară, s-ar putea spune că prima scală este mai validă. (Acest lucru se numește „validitate de criteriu").

Din păcate, în problemele legate de calitatea vieții, în general nu există criterii externe. O persoană poate simți că este nefericită, dar poate funcționa la un nivel înalt. Însăși ideea de calitate a vieții este subiectivă conceptual. Oricum a cui este calitatea vieții?

Prin urmare, de multe ori trebuie să ne bazăm pe validitatea conținutului, care este un amestec de bun simț și proprietăți psihometrice tehnice. Dacă vrem să știm dacă cineva se simte deprimat, am putea întreba: „Te simți foarte trist?" mai degrabă decât: „Te simți atletic?" Cu toate acestea, chiar și asta nu este atât de simplu, deoarece ceea ce cineva care nu este un expert în depresie ar putea considera în mod evident irelevant, cum ar fi tulburările de somn, este unul dintre cele mai puternice semne de depresie.

Desigur, dacă există un criteriu extern în raport cu care să se valideze o scală, acesta ar trebui utilizat. Dar chiar și validitatea conținutului poate deveni mai obiectivă, de exemplu, prin formarea unui grup de experți care să emită judecăți asupra validității conținutului articolelor. Pentru a testa acordul dintre judecători, poate fi utilizat coeficientul kappa, așa cum este descris în secțiune 3.3 .

7.4 Receptivitatea

Capacitatea de răspuns a unei scale este o măsură a cât de bine poate detecta schimbările ca răspuns la o intervenție. Reactivitatea sau sensibilitatea unei scale poate fi evaluată în mai multe moduri diferite și nu există un consens cu privire la care este cel mai bun. Unele concepte înrudite sunt descrise mai jos, care se referă la situația în care vă uitați la schimbarea măsurilor pre- și post-tratament.

(1) Utilizarea scorurilor de modificare (pre-post) este adecvată atunci când variabilitatea între pacienți este mai mare decât variabilitatea în interiorul pacienților. În general, scorurile de modificare pot fi utilizate în siguranță dacă

$$\sigma^2_{\text{între pacienți}} \wedge 05$$
$$\sigma_{\text{între}} \wedge \sigma_{\text{eroare}}$$

σ^2 între pacienți și eroarea σ^2 pot fi obținute dintr-o analiză a variantei scorurilor unui grup de pacienți care au replicat măsuri, astfel încât să puteți estima varianța datorată erorii.

(2) Un coeficient de sensibilitate la schimbare datorată unui tratament este

$$\sigma_{\text{schimbare}}$$
$$\sigma_{\text{schimba}} + \sigma_{\text{eroare}}$$

Pentru a obține $\sigma^2_{\text{eroarea}}$, trebuie să faceți o analiză a variantei măsurilor repetate pe aceleași subiecte. Sunt disponibile programe de calculator. Explicații detaliate ale acestui lucru apar în texte mai avansate.

(3) Mărimea efectului este pur și simplu modificarea scalei de la înainte la după tratament, împărțită la abaterea standard la momentul inițial. Abaterea standard este un indice al variabilității generale a scorurilor între grupul de persoane din studiu. Se poate măsura magnitudinea modificării medii a scorurilor după un anumit tratament, determinând ce procent din „variația de fundal" reprezintă această modificare . *Dimensiunea efectului* =

$$\frac{\textit{scorul mediu al schimbării}}{\textit{abaterea standard a scorurilor de bază (sau pretratament).}}$$

(4) O *măsură a receptivității* propusă de Guyatt și colab. [30] este

$$\frac{\textit{scorul mediu al schimbării}}{\textit{abaterea standard a scorurilor de schimbare pentru „ subiecții stabili ”}}$$

„Subii stabili” sunt greu de definit, dar ceea ce sugerează acest lucru este că poate fi utilizat un grup de control care nu primește intervenția sau primește placebo. Apoi se poate folosi abaterea standard a scorurilor de schimbare din grupul de control ca numitor în termenul de mai sus.

Variabilitatea scorurilor de schimbare în grupul de control (sau într-un grup de subiecți stabili) poate fi privită ca „variabilitatea de fundal” a *schimbărilor* , iar instrumentul de măsurare răspunde la gradul în care poate detecta schimbări dincolo de acest fond. variabilitate.

(5) Atunci când se evaluează schimbarea datorată tratamentului, ar trebui să existe *întotdeauna* un grup de control (adică, un grup fără intervenție sau placebo) pentru comparație, deoarece schimbarea poate apărea și la pacienții de control, iar întrebarea de interes este dacă înainte și după tratament. schimbarea în grupul de tratament depășește schimbarea „de bază ” în grupul de control. Dacă utilizați dimensiunea efectului ca măsură, atunci ar trebui să comparați dimensiunea efectului în grupul de tratament cu dimensiunea efectului în grupul de control.

Un exemplu numeric al acestor concepte este oferit în Anexa 6 .

7.5 Câteva capcane potențiale

(a) Multiplicitatea variabilelor

Cercetarea calității vieții se ocupă adesea de o mare cantitate de variabile. Să presupunem că un investigator examinează efectele unui medicament pentru tratarea hipertensiunii arteriale și îl compară cu placebo. Investigatorul poate avea câteva sute de elemente pentru a evalua diferite simptome fizice și psihologice și efecte secundare. Dacă s-ar compara cele două grupuri prin testul t pe fiecare dintre itemi, la nivelul de semnificație $p = 0{,}05$, ne-am

aștepta ca aproximativ 5 % din aceste teste să producă un rezultat semnificativ doar din întâmplare. Probabilitatea exactă este dificil de determinat, deoarece unele dintre aceste comparații ar fi corelate în virtutea faptului că aceiași pacienți răspund la toate, adică răspunsurile nu sunt independente. Dar, în orice caz, dacă anchetatorii aleg doar elementele semnificative și concluzionează că există efecte ale medicamentului, este posibil să comită erori de tip I, adică să respingă incorect ipoteza nulă.

De aceea, este important să folosiți scale care măsoară anumite constructe sau să grupați itemii într-un mod semnificativ din punct de vedere clinic. De exemplu, s-ar putea dori să măsoare depresia, anxietatea, ostilitatea și bunăstarea (fiecare constă din mai multe elemente). Pe de altă parte, anumite medicamente pot fi legate de simptome foarte specifice, cum ar fi coșmarurile, iar acest lucru ar putea trebui evaluat printr-un singur element care să întrebe despre frecvența coșmarurilor.

Ideea este că cercetarea privind calitatea vieții ar trebui, în general, să fie condusă de unele ipoteze specifice. În caz contrar, devine o „expediție de pescuit ” care doar pescuiește pentru orice important pe care îl poate găsi. Trebuie remarcat faptul că „expedițiile de pescuit” pot fi utile pentru a genera ipoteze care apoi trebuie testate într-un alt studiu.

(b) Generalizare

O altă problemă importantă este extrapolarea rezultatelor către alte populații decât cea din care a fost extras eșantionul de studiu. Efectele calității vieții pot fi diferite la bărbați decât la femei, la persoanele mai tinere decât la persoanele în vârstă și pot diferi în funcție de grupurile etnice și culturale. Ar trebui să fii atent în a face generalizări. În plus, psihometric

Proprietățile unei scale într-o limbă pot să nu fie aceleași ca într-o altă limbă, așa că cercetătorul trebuie să fie atent să se asigure că o scală tradusă are același sens ca scala originală. La construirea unei scale, trebuie consultat un psihometrician.

(c) Necesitatea unor standarde riguroase de cercetare

Unii oameni consideră că măsurile de calitate a vieții sunt „soft”. Ceea ce înseamnă în general este că ei cred că astfel de măsuri sunt subiective, variabile și poate lipsite de sens. Este un nonsens și, în măsura în care este

adevărat în unele studii, reflectă insuficiența cercetătorului, nu a subiectului. Aceste măsuri ar trebui să fie subiective din perspectiva pacientului , deoarece reflectă percepția subiectivă a pacientului despre bunăstare sau stres. Este cercetătorul care nu ar trebui să fie subiectiv și care nu trebuie să fie dacă urmează principiile cercetării. Variabilitatea măsurătorilor calității vieții nu este mai mare decât în multe măsuri fiziologice și, în orice caz, face parte din esența unor constructe ale calității vieții. În ceea ce privește sensul, aceasta este o problemă filozofică, nu științifică. Din punct de vedere științific, „sensul" ar trebui definit operațional. Cercetarea privind calitatea vieții ar trebui să respecte principiile oricărei cercetări bune, iar abordarea generală este aceeași ca pentru orice investigație științifică:

(1) Formulați o ipoteză testabilă.
(2) Cuantificați variabila (sau variabilele) dependentă.
(3) Selectați un design de studiu care poate răspunde la întrebarea pe care ați pus-o.
(4) Cuantificați variabilele independente.
(5) Controlul potențialilor factori de confuzie (prin proiectarea studiului și/sau analiza datelor).
(6) Planificați o dimensiune a eșantionului care vă va oferi suficientă putere pentru a detecta dimensiunea efectului de interes.
(7) Încercați să vă asigurați că minimizați părtinirea sistematică.
(8) Colectați datele, acordând multă atenție controlului calității.
(9) Analizați datele folosind tehnici statistice adecvate.
(10) Faceți inferențe în concordanță cu punctele forte și limitările studiului.

Chapter 8
MAI MULT DESPRE EPIDEMIOLOGIA GENETICĂ

> Să presupunem atunci că mintea este, după cum spunem, hârtie albă (tabula rasa), lipsită de toate caracterele fără nicio idee; cum se face sa fie mobilat? De unde provine acel magazin vast, pe care fantezia ocupată și nemărginită a omului l-a pictat pe el într-o varietate aproape nesfârșită? ..... La aceasta răspund, într-un cuvânt, Din experiență: în aceea că toată cunoștințele noastre se întemeiază .....
>
> John Locke
> Un eseu despre înțelegerea umană (1689)

8.1 O nouă eră științifică

Suntem departe de a crede că mintea este o „tabula rasa", o tablă goală. Știm acum că multe sunt de fapt înnăscute, adică sub influență genetică. Scopul acestui capitol este de a-i ajuta pe cei care doresc să citească literatura în expansiune rapidă în epidemiologia genetică. Astfel, este o privire de ansamblu asupra proiectelor de bază și a statisticilor utilizate în acest domeniu; nu este cuprinzător și nici nu este foarte tehnic.

Accentul cercetării epidemiologice a evoluat pe măsură ce s-au înregistrat progrese paralele în alte domenii ale medicinei și științei fundamentale. În epoca în care bolile infecțioase erau rampante, epidemiologia s-a preocupat de identificarea surselor de infecție și a metodelor de transmitere, în mare parte prin munca de teren. Pe măsură ce agenții infecțioși au fost descoperiți, odată cu îmbunătățirea stării sanitare și a stării de sănătate, bolile cronice, cum ar fi bolile de inimă și cancerul, au devenit principalele cauze de deces și dizabilități în lumea dezvoltată și au ajuns să fie ținta principală a cercetării epidemiologice. (Acum că noi boli infecțioase apar din nou, această parte a epidemiologiei câștigă din nou proeminență).

Obiectivul epidemiologiei bolilor cronice a fost identificarea factorilor de risc pentru aceste boli. Această parte a poveștii a avut un mare succes de sănătate publică. Acum știm, datorită studiilor epidemiologice, care sunt factorii de risc majori modificabili pentru bolile cardiovasculare: hipertensiunea arterială,

colesterol ridicat și LDL, fumat, exces de greutate și inactivitate. Provocarea noastră acum este să găsim modalități de a face schimbări în stilul de viață al populației, ceea ce va reduce și mai mult ratele bolilor cardiovasculare. De asemenea, cunoaștem multe dintre expunerile legate de cancer, dar nu la fel de cuprinzător ca pentru bolile de inimă.

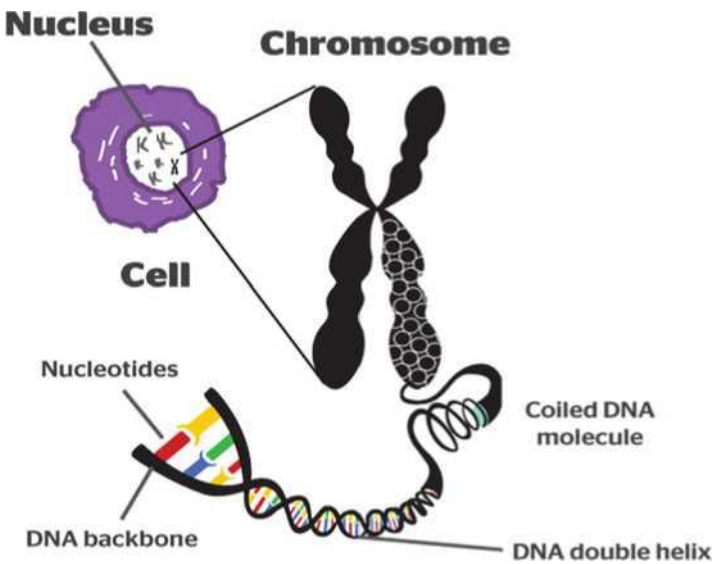
În acest moment istoric din punct de vedere științific, pe măsură ce știința intră pe deplin în era genomică, epigenomică și proteomică (și alte „ omice ”), - epidemiologia a intrat într-o nouă fază a activității de cercetare: epidemiologia moleculară . Aceasta este căutarea de biomarkeri din sânge sau țesut și polimorfisme genetice (variante) care sunt asociate sau predispun la boli. De ce este acesta diferit de orice alt factor de risc investigat în epidemiologie ? În multe privințe, nu este , mai ales în ceea ce privește biomarkerii din sânge, dar în epidemiologia genetică, există modele de studiu și metode de analiză statistică care sunt destul de diferite. Un aspect cu adevărat nou al epidemiologiei moleculare și genetice este adevărata colaborare a oamenilor de știință de bază, a medicilor și a epidemiologilor. De prea mult timp disciplinele au urmat căi de cercetare separate, iar oamenii de știință citesc mai ales reviste științifice din domeniul lor. Dar epidemiologia moleculară nu poate continua cu succes fără interfața oamenilor de știință de laborator și a cercetătorilor populației.

Mai jos sunt câteva elemente de bază ale geneticii pe care le puteți sări peste citire dacă toate acestea sunt familiare. ADN-ul (acidul dezoxiribonucleic) este format din patru unități sau nucleotide. Aceste nucleotide, numite și baze, sunt adenină, guanină, timină și citozină și sunt notate cu literele A, G, T și C. ADN-ul este aranjat în două catene răsucite într-o formă de dublu helix, astfel încât nucleotidele AGCT împerechea între ele în moduri fixe. Un A se perechează întotdeauna cu T și C se face întotdeauna perechi cu G. Acestea se numesc perechi de baze. Dacă un fir din elica dublă ar fi înșirat într-o linie, ar putea arăta astfel:

AATTCGTCAGTCCC. Cealaltă componentă care se asociază cu ea ar fi TTAAGCAGTCAGGG.

Există trei miliarde de perechi de baze (sau șase miliarde de baze) în genomul uman (care se referă la tot materialul genetic la om). Aceste trei miliarde de perechi de baze sunt organizate în 23 de perechi de cromozomi (una de la mamă și una de la tată), care se află în fiecare celulă vii din organism (cu excepția spermatozoizilor și a ovulelor, fiecare dintre ele având câte un cromozom.

până când se unesc pentru a forma o celulă fertilizată care are acum întregul cromozomi). În aceste trei miliarde de perechi de baze, există aproximativ 20.000 de gene care sunt secvențe de perechi de baze de lungimi diferite și care oferă codul pentru formarea proteinelor. Secvențele rămase servesc în principal funcții de reglare sau au funcții care sunt necunoscute în prezent. Cele mai frecvente variații ale genomului sunt cunoscute ca polimorfisme cu un singur nucleotide (SNP, pronunțate ca „snips") și implică o diferență într-o singură literă a codului genetic. Unele SNP sunt variante normale într-o populație, unele pot proteja împotriva bolilor, iar altele pot predispune la boli.



8.2 Prezentare generală a epidemiologiei genetice

Epidemiologia genetică urmărește să identifice genele legate de boală și să evalueze impactul factorilor genetici asupra sănătății și bolii populației. Iată o prezentare generală a strategiei adesea folosite pentru a studia determinanții genetici ai bolii. Mai întâi putem dori să stabilim dacă boala se transmite în familii . Dacă este

nu este familial, nu este probabil să fie moștenit; dacă este familială, poate fi sau nu din cauza unor factori genetici (mediile se desfășoară și în familii). Apoi, vrem să vedem dacă variația genetică contribuie la transmiterea familială. O metodă pentru a determina acest lucru este studierea gemenilor (descris în secțiunea 8.3). Dacă stabilim că boala este ereditară , am dori să localizăm și să identificăm genele implicate. Și, în sfârșit, dorim să înțelegem mecanismele prin care aceste gene contribuie la apariția bolii.

Ca prim pas, putem dori să aflăm unde sunt localizate genele care contribuie la tulburare. O abordare în acest sens este de a efectua studii de legătură între indivizii afectați de boală și familiile acestora (descrise în secțiunea 8.4). Studiile de legare pot identifica regiuni de pe cromozom care sunt susceptibile de a adăposti genele bolii. Odată ce am identificat una sau mai multe astfel de regiuni, putem căuta să vedem ce gene se știe că locuiesc în acele regiuni. Putem apoi testa aceste gene folosind studii de asociere la indivizi neînrudiți (descrise în secțiunea 8.6) pentru a determina dacă vreo variantă (numită și alele) ale acestor gene este asociată cu boala. O altă abordare, mai potrivită pentru bolile complexe comune, este de a începe cu studii de asociere la nivel de genom (discutate în secțiunea 8.10).

Deci, există o varietate de modele și teste statistice care pot fi utilizate pentru a defini baza genetică a unei boli, inclusiv (1) studii pe gemeni pentru a determina dacă boala are o componentă ereditară; (2) studii de legătură pentru identificarea și localizarea regiunilor cromozomilor care conțin gene implicate în boală; și (3) studii de asociere pentru a determina dacă anumite variante genetice sunt asociate cu boala, pentru a examina modul în care acestea interacționează cu mediul și pentru a determina modul în care afectează sănătatea populației. Vom limita discutia la cateva modele destul de simple care vor da savoarea subiectului. Cititorii interesați de mai multă profunzime pot consulta multe mai multe scrieri tehnice pe acest subiect, enumerate în secțiunea de referință [31–35] pentru acest capitol.

8.3 Studii pe gemeni

Pentru a explora o influență genetică asupra bolii, putem mai întâi să vedem dacă aceasta se regăsește în familii. Dar ceva care este familial nu este neapărat moștenit. De exemplu: părinții obezi au copii obezi din cauza geneticii sau din cauza nutriției și a nivelurilor de activitate care se transmit de la părinți la copii? Ceea ce vrem să știm este dacă și în ce măsură fenotipul (ceea ce observăm la persoană, de exemplu, obezitatea) este afectat de factori genetici.

O modalitate de a evalua influența variației genetice este din studiile pe gemeni. Gemenii identici (monozigoți – care provin din același ovul fertilizat) împărtășesc 100 % din genele lor, în timp ce gemenii fraterni (dizigoți – care provin din două ouă fertilizate) au în medie 50 %, la fel cum fac frații negemeni. O modalitate de a estima puterea influențelor genetice este de a calcula ereditabilitatea , h^2. Pentru studiile pe gemeni, ereditabilitatea poate fi calculată ca de două ori diferența dintre corelația pentru acea trăsătură între gemenii monozigoți minus corelația la gemenii dizigoți sau

$$h — ^2 (^r mz \, ^r dz)$$

Luați în considerare tensiunea arterială. Dacă variația stării sau trăsăturii investigate ar fi complet atribuibile variației genetice, atunci fiecare membru al unei perechi de gemeni monozigoți ar fi afectat în mod egal (fiecare membru ar avea aceeași tensiune arterială) și corelația dintre gemenii monozigoți ar fi 1,0; corelația la gemenii dizigoți ar fi însă .50.

În acest caz, h^2 ar fi 2(1-0,5) — 1,0 sau 100 %. Dacă condiția nu este complet ereditară în populație, atunci $r_{mz} — r_{dz}$ și $h^2 — 0$. Deoarece bolile și trăsăturile sunt în general parțial ereditare, h^2 se află undeva între 0 și 1,0.

Dacă vorbim de variabile continue, ne putem gândi la heritabilitate în termeni de coeficienți de corelație. Dacă vorbim de variabile categorice, putem vorbi de rate de concordanță, unde

$$h^2 \quad \frac{\% \textit{ gemeni monozigoti concordantipentru boala} — \% \textit{ Cfdizigoticgemeni concordanti}}{1 — \% \textit{ gemeni dizigotici concordanţi}}$$

Unele estimări aproximative au raportat [31–33] de ereditabilitate din studiile pe gemeni sunt .60 pentru alcoolism, .30–.50 pentru trăsăturile de personalitate, .35 pentru cancerul colorectal, .26 pentru scleroza multiplă, 0.75 pentru înălțime și .80 pentru schizofrenie.

Este important să ne amintim că ereditatea nu măsoară cât de mult din boala unui individ este atribuită geneticii; mai degrabă ne spune ce proporție din variabilitatea populației în fenotip este rezultatul variației genelor din populație. Deci este o măsură aplicabilă unei populații, nu unui individ. Dacă aveți oameni care trăiesc exact în același mediu, atunci orice variație pe care o întâlniți în fenotip s-ar datora în principal factorilor genetici, deoarece nu există nicio variație de mediu. Într-un astfel de caz, dacă toți factorii de mediu sunt constanți pentru

populație, ereditabilitatea ar fi de 100 %. Deci, există unele limitări ale acestei măsuri, dar ne oferă o idee în ce măsură variația genetică contribuie la variația fenotipică într-o populație. Cu toate acestea, ereditatea nu ne spune nimic despre genele care sunt responsabile pentru acea variație, care variante genetice sunt implicate, câte variante sunt implicate sau care sunt dimensiunile efectului lor. Aceste informații mai detaliate sunt denumite „arhitectura genetică" a unei trăsături sau boli.

8.4 Studii de legături și asocieri

Dacă știm că o boală este ereditară, ne putem întoarce acum la sarcina de a identifica efectiv genele implicate. Cele mai multe tulburări care sunt studiate de epidemiologi (de exemplu, boli cardiovasculare, tulburări psihice , forme comune de cancer) sunt considerate tulburări „complexe". Spre deosebire de tulburările cu o singură genă sau mendeliană, cum ar fi fibroza chistică sau boala Huntington , se crede că bolile complexe rezultă din contribuția mai multor sau mai multe gene care interacționează cu factorii de risc de mediu. Acest lucru poate face identificarea efectului unei gene individuale o sarcină destul de dificilă. Efectul unei anumite alele în cadrul acelei gene

poate fi destul de mic. Este un pic ca și cum ai căuta proverbialul ac în carul de fân. Cu toate acestea, genele care contribuie la boli sunt descoperite și există anumite strategii care sunt folosite în căutare.

Unde în genom căutăm genele care conferă susceptibilitate la boală? O modalitate de a răspunde la această întrebare este utilizarea analizei legăturilor genetice .

(a) Analiza legăturii se bazează pe fenomenele de încrucișare și recombinare care apar în timpul procesului de meioză când se formează celulele sexuale (sperma și ovulul). Fiecare persoană are două copii ale fiecăruia dintre cei 23 de cromozomi care alcătuiesc genomul: o copie este moștenită de la mamă și una de la tată. În timpul formării spermatozoidului și a ovulelor, aceste 23 de cromo, unele perechi se aliniază și schimbă segmente de material genetic într-un proces cunoscut sub numele de crossing over . Această recombinare are loc în

unul sau mai multe locuri de-a lungul cromozomului. Cu cât doi loci sunt mai apropiați pe un cromozom, cu atât este mai puțin probabil să apară un eveniment de recombinare între ei și, prin urmare, cu atât este mai probabil ca aceștia să fie moșteniți împreună . Loci care tind să fie co-moșteniți se spune că sunt legați genetic. Putem folosi acest fapt pentru a estima distanța dintre doi loci sau markeri genetici (un marker genetic este o variație ADN a cărei localizare cromozomială este cunoscută). Distanța fizică dintre doi markeri este invers legată de cât de frecvent sunt moșteniți în comun de-a lungul generațiilor într-o familie.

(b) Distanța dintre doi loci se măsoară uneori în centi morgani . Un centimorgan (cM) este o unitate de distanță de-a lungul unui cromozom, dar nu în sensul obișnuit al distanței fizice. Este într-adevăr o măsură de probabilitate care este o reflectare a distanței fizice; reflectă probabilitatea ca doi markeri sau loci să fie separați (sau segregați) prin încrucișare în timpul meiozei. Dacă cei doi markeri sunt foarte apropiați unul de celălalt, nu se vor separa (spunem că sunt „legați"); dacă sunt departe unul de celălalt, este mai probabil să se încrucișeze și materialul genetic este recombinat în timpul meiozei . Apoi, acest ADN recombinat este transmis descendenților. Doi loci sunt la un centimorgan separat dacă

probabilitatea ca acestea să fie separate prin încrucișare este de numai 1 % (o dată la o sută de meioze). S-a estimat că există aproximativ 1 milion de perechi de baze într-un interval de 1 cM. Locii care sunt departe unul de celălalt, să zicem 50 cM, vor fi moșteniți independent unul de celălalt, așa cum ar fi dacă ar fi pe cromozomi diferiți. Scopul studiilor de legătură este de a localiza gena de susceptibilitate a bolii pentru a se afla într-o anumită regiune a cromozomului .

(c) Așadar, am putea începe căutarea unei gene a bolii prin colectarea familiilor afectate de boală și efectuând o analiză a legăturii folosind markeri distanțați la intervale (să zicem la 10 cM unul de celălalt) de-a lungul întregului genom. Dacă găsim un marker care pare a fi co-transmis cu boala, am avea dovezi că markerul este legat genetic de o genă a bolii. Cu alte cuvinte, este probabil să existe o genă pentru boală în aceeași regiune cu markerul legat.

(d) După ce am găsit o regiune cromozomială legată de boală, am putea încerca să restrângem regiunea prin genotiparea și testarea markerilor

suplimentari în acea regiune (să zicem la intervale de 1 cM). Cu toate acestea, chiar și această regiune relativ mică poate conține multe gene.

(e) Următorul nostru pas ar putea fi să analizăm genele despre care se știe că locuiesc în această regiune. Ne-ar interesa în special genele care au o legătură plauzibilă cu boala de interes (acestea ar fi „ gene candidate ”) bune. De exemplu, dacă studiem diabetul, genele care produc proteine implicate în metabolismul glucozei ar fi „gene candidate” importante.

(f) Acum putem vedea dacă anumite alele (variante) ale genelor din acea regiune cromozomială sunt asociate cu boala. Acest lucru se poate face prin:

(1) Studii de asociere care utilizează metode caz-control la persoane neînrudite , examinând dacă o alelă este mai frecventă în cazuri decât martori (descris în Sect. 8,6)

(2) Studii de asociere în familii pentru a vedea dacă o alelă este transmisă mai frecvent la cazuri decât se aștepta întâmplător (descris în Sect. 8,9)

Deci, în esență, analiza legăturii ne spune că o anumită locație a markerului este în apropierea unei gene de susceptibilitate a bolii; analiza de asociere ne spune că o anumită alelă a unei gene sau a unui marker este mai frecvent moștenită de persoanele cu boală.

8.5 Scor LOD: Statistică de legătură

Statistica clasică utilizată pentru a evalua puterea dovezilor în favoarea legăturii unui marker genetic și a genei de susceptibilitate a bolii este scorul LOD ($\log_{10}$ al șanselor în favoarea legăturii). Acesta va fi descris doar în principiu, pentru a ajuta la interpretarea articolelor epidemiologice care se ocupă de genetică. Calculele reale sunt complexe și necesită pachete speciale de programe.

Principiul care stă la baza scorului LOD este descris în secțiunea anterioară: dacă avem doi loci - să zicem, un marker și o genă a bolii - cu cât sunt mai aproape de un cromozom, cu atât este mai mică probabilitatea ca ei să fie separați de un eveniment de recombinare în timpul meiozei și cu atât este mai probabil ca acestea să fie co-moștenite de descendenți.

Probabilitatea recombinării, numită fracțiune de recombinare, este notă cu

simbolul θ și depinde de distanța dintre genă și marker. Dacă nu există recombinare și gena și markerul sunt complet legate, atunci fracția de recombinare este 0 . Valoarea maximă a lui θ este .5 (dacă gena și markerul au fost moștenite independent, atunci probabilitatea ca markerul să fi fost transmis, dar nu gena = probabilitatea ca gena să fi fost transmisă, dar nu markerul = .50).

Deci, dacă doriți să știți dacă există o legătură, trebuie să estimăm cât de probabil este ca θ să fie mai mic de .5, având în vedere datele pe care le-am observat. Pentru aceasta folosim raportul de probabilitate, care, după cum vă amintiți din capitolul 2 este raportul dintre probabilitatea simptomelor observate, dată boală împărțită la probabilitatea simptomelor observate în cazul în care nu există boală. În acest caz

$$LR = \frac{\text{Probability observed heritance data , data linking}}{\text{Probabilitatea datelor de moștenire observate , având în vedere nicio legătură}}$$

Ipoteza nulă aici nu este nicio legătură (sau fracțiune de recombinare θ = .5), iar ipoteza alternativă este legătura (sau θ < .5). Dacă respingem nulul, „acceptăm" ipoteza alternativă. Statisticile de testare utilizate pentru a vedea dacă avem date suficiente pentru a concluziona legătura este scorul LOD care este $\log_{10}(LR)$. Pentru tulburările mendeliane (cu o singură genă), un scor LOD de 3 a fost în mod tradițional pragul pentru declararea unei legături semnificative, deși pentru tulburările complexe au fost recomandate praguri mai mari (3,3–3,6). Un scor LOD de 3,0 indică 10^3 cote în favoarea legăturii în comparație cu nicio legătură, adică 1.000:1 cote în favoarea legăturii.

Din motive complexe dincolo de scopul acestei cărți (dar descrise în referințele de la sfârșit), un scor LOD poate fi tradus în probabilitate prin înmulțirea lui cu constanta 4,6 : LOD × 4,6 este distribuit ca chi-pătrat cu 1 grad de libertate. . (4,6 este de 2 ori logul natural de 10.) Astfel, un LOD de 3,0 este echivalent cu un chi-pătrat de 3 × 4,6 = 13,82 și corespunde cu p = .0002. Datele de moștenire pentru analizele de legături pot proveni din studii de genealogie a familiei, de la frați sau alte grupuri de familie.

Analiza de legare a scorului LOD este uneori denumită analiză de legare „parametrică", deoarece necesită specificarea anumitor parametri (de exemplu, frecvența alelelor de boală și marker, modul recesiv vs. dominant de moștenire, penetranța genei bolii). Atunci când acești parametri sunt cunoscuți sau pot fi

aproximați, analiza parametrică a scorului LOD este cea mai puternică metodă de analiză a legăturilor. Acest lucru poate fi adevărat pentru boala mendeliană (cu o singură genă), dar pentru multe tulburări complexe, acești parametri nu sunt cunoscuți. Metodele de legare „neparametrice" (cunoscute ca abordarea partajării alelelor) sunt adesea folosite pentru a studia tulburările complexe, deoarece nu necesită cunoașterea modului de moștenire sau a altor parametri genetici. Există o serie de statistici disponibile, descrise în textele mai avansate.

8.6 Studii de Asociere

În comparație cu analiza legăturii, studiile de asociere sunt mai strâns asemănătoare cu studiile epidemiologice tradiționale și se bazează cel mai adesea pe designul caz-control. Într-un studiu de asociere, anchetatorii sunt interesați să descopere

dacă există vreo asociere între o anumită alele la un locus polimorf și fenotipul în cauză. (Un polimorfism este o variație a secvenței ADN care apare la cel puțin 1 % din populație. O variație care apare la mai puțin de 1 % este denumită mutație). În scopul acestei discuții, vom presupune că polimorfismele pe care le analizăm sunt SNP-uri (polimorfisme cu un singur nucleotide) sau variante într-o singură dintre bazele A, T, C, G (reprezentând adenină, timină, guanină, citozină).) la un anumit loc. Rețineți că există multe alte clase de variații ADN, inclusiv inserții și deleții mici de nucleotide și variații ale numărului de copii care implică ștergerea sau duplicarea unor bucăți mai mari de ADN, așa cum este ilustrat mai jos.

Copiați numărul Variantă

ATTGCKXITCGGGGGTZATIA

ATTGGCCTTA ..... -

Retipărit cu permisiunea de la Macmillan Publishers Ltd: [Nature Reviews Genetics] (Frazer KA, Murray SS, Schork NJ, Topol EJ, Human genetic variation and its contribution to complex features), copyright (2009)

Deci, să spunem că la un anumit SNP unii oameni au alela A și alții au alela G. Vrem să știm dacă persoanele cu boală au mai multe șanse să spună, alela A decât alela G. Pentru un fenotip binar (de exemplu, bolnav sau nu), putem face studii caz-control de asociere, luând cazuri care sunt afectate de boală și controale neînrudite care nu sunt. Amintiți-vă că fiecare persoană primește o copie a unei alele la un anumit loc de la mamă și o copie

de la tată. (Aceștia se află exact în același loc pe fiecare cromozom al cromozomilor perechi). Deci, dacă o persoană primește un A de la mamă și un G de la tată, genotipul persoanei respective este AG. Dacă persoana primește un A de la fiecare părinte, genotipul persoanei respective este AA. Putem compara frecvențele genotipurilor AA, AG sau GG între cazuri și martori, reprezentate de numerele 0, 1 sau 2 pentru numărul de alele minore pe care le conține (o alela majoră este cea mai comună în populație, o alela minoră este cea mai puțin obișnuită). Deci, în exemplul nostru, să presupunem că alela G este alela minoră; apoi putem converti genotipurile în numere după cum urmează: 0 pentru genotipul AA, 1 pentru genotipul AG (deoarece conține un G) și 2 pentru genotipul GG. Apoi putem vedea dacă numărul de alele minore (0, 1 sau 2) diferă între cazuri și controale. Putem folosi teste statistice obișnuite ale diferențelor dintre proporții sau regresii logistice multiple (vezi Secțiunea 4.17) pentru a determina raportul de șanse conectat cu alela în cauză și putem testa interacțiunile genă-mediu prin includerea unui termen de interacțiune al prezenței. ale alelei și un anumit factor de mediu, cum ar fi fumatul. Pentru un fenotip continuu (de exemplu, tensiunea arterială), putem folosi regresia liniară.

Studiile de asociere pot fi mai puternice decât analiza legăturii pentru detectarea genelor cu efect modest, făcându-le o abordare atractivă pentru studierea tulburărilor complexe, care implică multe variante de risc cu efect individual relativ mic. Calculele de putere pentru testele de asociere pot fi efectuate

folosind mai multe instrumente online, inclusiv Calculatorul de putere genetică (http://pngu.mgh.harvard.edu/~purcell/gpc/).

8.7 Studii ale asociației de gene candidate

Studiile de asociere pot fi utilizate pentru a evalua o genă candidată, adică o genă pe care datele anterioare, adesea din studii de legătură sau studii biologice anterioare, au sugerat că este implicată în fenotipul nostru de interes. Cu toate acestea, cunoștințele noastre despre baza biologică a multor boli sunt incomplete. Astfel, probabilitatea anterioară ca orice genă candidată sau polimorfism să fie implicată de fapt este adesea scăzută. În această circumstanță, o asociere semnificativă este probabil să se datoreze întâmplării. Într-adevăr, descoperirile genelor candidate
au fost notoriu de dificil de replicat, ceea ce sugerează că multe asociații raportate sunt false. Ca rezultat, studiile genelor candidate au căzut în mare parte în dizgrație și au fost înlocuite cu analize la nivel de genom (GWAS, discutat în secțiunea 8.10) care nu se bazează pe ipoteze anterioare.

8.8 Stratificarea populației sau Structura populației

O problemă importantă pentru studiile de asociere este alegerea grupurilor de control. Frecvența alelelor într-o populație poate diferi în funcție de ascendență din motive care nu au legătură cu boala de interes. Asemenea diferențe de fond genetic pot confunda rezultatele noastre din cauza unui fenomen denumit „ stratificarea populației ”. [36]

De exemplu, să presupunem că varianta A este mai frecventă la indivizii de origine europeană și varianta G este mai frecventă la cei de origine africană. Dacă se întâmplă ca grupul nostru de boală (caz) să aibă mai mulți europeni-americani și grupul nostru fără boală (de control) să aibă mai mulți afro-americani, atunci am putea descoperi că alela A este mai frecventă printre cei cu boală decât la cei fără boală, dar s-ar putea să fie doar o reflectare a faptului că am avut mai multe cazuri euro-americane decât controale euro-americane și că este mai probabil ca europeni-americanii să fie purtători ai alelei A. Deci, am avea o constatare fals

pozitivă din cauza ascendenței fără egal a celor două grupuri.

Exemplul numeric de mai jos ilustrează acest punct. Puteți sări peste citirea exemplului dacă doriți, dar acesta este afișat aici pentru cei care doresc să-l lucreze numeric. Să presupunem că 80 % dintre persoanele de ascendență europeană poartă alela A la un anumit SNP și doar 40 % dintre persoanele de ascendență africană poartă alela A. Procentele respective pentru alela G sunt de 20 % la europeni și de 60 % la ascendența africană (Tabelul). 8.1). Acum să spunem că se întâmplă ca dintre cazurile de boală, 70% să fie de origine europeană, dar în controale doar 10% sunt de origine europeană (Tabel 8.2). La ce % din alelele A ne-am aștepta în cazuri și controale?

Ei bine, dintre cazuri, deoarece 70 % sunt de origine europeană și 80 % dintre aceștia poartă alela A, ne-am aștepta la .70 × .80 = .56 sau 56 % din

Tabelul 8.1 Frecvențele alelelor unui SNP ipotetic în funcție de ascendența populației

Grup etnic	european-american	afro-american
Prevalența alelei A	80,0 %	40,0 %
Prevalența alelei G	20,0 %	60,0 %

Tabelul 8.2 Prevalența bolii ipotetice în funcție de ascendența populației

Cazuri de boală		Controal
% european-american	70,0 %	10,0 %
% afro-americani	30,0 %	90,0 %

Tabelul 8.3 Calcule

O alela	Cazuri	Controale
EA	.70 × .80 = .56	.10 × .80 = .08
AA	.30 × .40 = .12	.90 × .40 = .36
Alela totală A	.56+ .12 = .68	.08 + .36 = .42
alela G	Cazuri	Controale
EA	.70 × .20 = .14	.10 × .20 = .02
AA	.30 × .60 = .18	.90 × .60 = .54
Alela G totală	.14 +.18 = .32	.02 +.56 = .58
Rezumatul acestui tabel		
	Cazuri	Controale
Prevalența alelei A	68,0 %	42,0 %
Prevalența alelei G	32,0 %	58,0 %

cazuri de ascendență europeană să poarte alela A; restul de 30% dintre cazuri sunt de ascendență africană la care frecvența alelei A este de 40%, așa că ne-am aștepta la încă .30 × .40 = .12 sau 12% să poarte alela A, pentru un total de 56% + 12 % = 68 % din cazuri purtând alela A, așa cum se arată în tabel 8.3 .

Tabelul 8.4 Exemplu de studiu caz-control cu 1.000 de cazuri și 1.000 de controale

	Cazuri		de control	
Prevalence of A allele	680	420	1.100	
Prevalence of G allele		320		580 900
	1.000		1.000	2.000

În controale am avea doar 10 % europeni dintre care 80 % poartă alela A, așa că ne-am aștepta ca 0,10 × 0,80 = 0,08 sau 8 % din controalele strămoșilor europeni să poarte alela A; 90% dintre controalele noastre sunt de origine africană, pentru care frecvența alelei A este de 40%, așa că ne așteptăm ca .90 × .40 = .36 sau 36% dintre ei să poarte alela A. Combinând controalele ancestrale europene și africane, ne așteptăm ca 8 % + 36 % = 42 % să poarte alela A. Se fac calcule similare pentru alela G, așa cum se arată în rezumatul tabelului 8.3 .

Acum să presupunem că avem un studiu de 1.000 de cazuri și 1.000 de controale și nu există o asociere reală între alela A și boala noastră. Cu toate acestea, așa cum se arată în tabel 8.4 , frecvențele alelelor bazate pur și simplu pe diferențele populației din eșantionul nostru caz-martor ar indica o asociere (chi-pătrat = 136,6; $p < 0{,}00001$). Așa că trebuie să luăm în considerare foarte atent stratificarea populației.

Există mai multe opțiuni disponibile pentru a rezolva problema confuziei prin stratificarea populației. În primul rând, s-ar putea încerca să restricționeze analiza la cazuri și controale cu fond genetic omogen – de exemplu, cele de ascendență europeană. Acest lucru se face frecvent prin limitarea eșantionului care urmează să fie studiat pe baza etniei auto-raportate. Cu toate acestea, studiile genetice ale populației au arătat că subgrupurile populației pot exista chiar și în cadrul unor categorii largi ca aceasta, astfel încât confuzie reziduală pot fi prezente.

De asemenea, se poate aplica o tehnică cunoscută sub numele de „ control genomic ” care se ajustează pentru un factor de inflație numit λ (lambda). Pentru a face acest lucru, genotipăm (adică determinăm alelele pentru) un set de 50 sau mai multe SNP-uri din genom care se presupune că nu au legătură cu boala de interes (SNP-uri „nule”). Apoi calculăm statisticile testului (chi-pătrat) comparând cazuri și

controale pentru SNP-urile noastre nule. Acum obținem mediana acestor valori chi-pătrat observate pentru SNP-urile noastre nule. Mediana așteptată a unei distribuții chi-pătrat cu un grad de libertate unde într-adevăr nu există nicio diferență (adică, distribuția nulă) se dovedește a fi 0,456. Calculăm λ ca mediana

statisticilor chi-pătrat observate pentru markerii nuli împărțiți la 0,456. Acum putem corecta chi-pătratele noastre observate pentru SNP-urile candidate, împărțindu-le la λ .

În cele din urmă, putem evita problema stratificării populației utilizând un design bazat pe familie, mai degrabă decât caz-control, așa cum este descris în secțiunea 8.9 .

Astfel, dacă găsim o asociere între o anumită alelă și boala pe care o studiem, aceasta poate fi din unul dintre cele patru motive: (1) ar putea fi un fals pozitiv din cauza întâmplării (eroare de tip I); (2) ar putea fi un fals pozitiv din cauza confuziei din cauza stratificării populației; (3) poate fi în „dezechilibru de legătură" cu alela adevărată a bolii, ceea ce înseamnă că alela pe care am găsit-o mai frecvent în cazuri decât la martori este situată fizic suficient de aproape de alela adevărată a bolii încât cele două alele tind să fie moștenite împreună și co-apare la indivizii afectați; și (4) există într-adevăr o adevărată asociere cauzală a alelei pe care am studiat-o și a bolii. (Așa cum am spus mai devreme, genetica și viața nu sunt simple).

8.9 Asociația bazată pe familie și testul de dezechilibru al transmisiei (TDT)

Un test statistic al legăturii și al asocierii este testul de dezechilibru al transmisiei . Acesta este un test de asociere bazat pe familie care evită confuzia din cauza stratificării populației prin examinarea transmiterii alelelor în cadrul familiilor. Dezechilibrul de transmisie se referă la o abatere de la legea mendeliană a transmiterii independente a alelelor la un loc dat. Să luăm din nou în considerare un SNP la care există două alele posibile, A și G. Un părinte care este heterozigot la acest SNP (adică are genotipul AG) poate transmite fie un A sau un G copilului. În ipoteza nulă a lipsei de legare și a asocierii, probabilitatea de a transmite oricare dintre aceste alele este de 50 %. Dacă observăm transmiterea alelei A mult mai des decât așteptările întâmplătoare (adică, mai mult

mai mult de 50 % din timp) la descendentul afectat, atunci concluzionăm că alela A este asociată cu boala.

Statisticile de bază sunt destul de simple. Unitatea de analiză este transmisia unei alele . Iată un exemplu folosind triouri. Un trio este un descendent afectat și ambii părinți. Există două transmisii posibile ale unei alele într-un anumit locus - una de la mamă și una de la tată. În diagrama de mai jos construim un tabel 2 × 2 și numărăm numărul de alele transmise care aparțin fiecărei celule a tabelului.

Trio 1: deoarece copilul afectat are două alele AA în locusul luat în considerare, ea trebuia să fi primit un A de la fiecare părinte; astfel, vedem că tatăl și-a transmis alela A și nu G și mama și-a transmis și alela A și nu G, indicând că aceste două transmisii aparțin celulei b care descrie transmiterea unei alele A și nu a unei G.

Alele transmise

AG AG	*Not transmitted*	Allele G	Allele A
AA	Allele G	Cell a	Cell b = 2
	Allele A	Cell c	Cell d

Trio 2: Aici o transmisie a fost a unui A și nu a unui G, iar cealaltă a fost a unui G și nu a unui A. (Nu știm care a fost de la ce părinte, dar știm că 1 transmisie este descrisă de celula b și altele prin celula c.)

		Transmitted	*Allele*
AG AG	*Not transmitted*	Allele G	Allele A
AA	Not G		1 (mother **or** father)
	Not A	1 (mother **or** father)	

Trio 3: Aici, A din copil trebuia să vină de la tată, deoarece asta era tot ce trebuia să transmită, așa că a transmis o alele A și nici nu și-a transmis cealaltă alele A; astfel, acea transmisie aparține celulei

d. Mama și-a transmis alela G și nu alela A și astfel că transmiterea aparține celulei c.

Alele transmise

AG AG *Not transmitted*

	Allele G	Allele A
AA Not G		
Not A	1 (mother)	1 (father)

Imaginează-ți că avem 120 de astfel de triouri. Acum vom combina datele din cele 240 de transmisii dintre cele 120 de triouri adăugând numerele din fiecare celulă într-un tabel rezumativ, așa cum se arată în exemplul de mai jos:

Alele transmise

Netransmis

	Allele G	Allele A
Nu G	65	90
Nu A	50	45

și calculați statistica de testare TDT ca cantitate:

$$\frac{(b - c)^2}{(b + c)} \qquad \frac{(90 - 50)^2}{140} \qquad 11{:}43$$

care este distribuit ca x^2, cu un grad de libertate și, deoarece este mai mare de 3,84, putem respinge ipoteza lipsei de legătură la un nivel alfa de 0,05 (deși, așa cum vom discuta, acest prag statistic este de obicei considerat și el liberale pentru studii de asociere genetică). TDT-ul este într-adevăr un test al lui McNemar (descris în Sect. 3.2), analog analizei caz-control potrivite, unde controalele sunt alele netransmise, mai degrabă decât persoane.

Rețineți că orice transmisii care ajung în celula a sau d sunt neinformative. Statistica de testare TDT utilizează numai informații din celulele b și c. Astfel, doar părinții care sunt heterozigoți (având un A și un G) sunt informativi. De când începem cu copiii afectați identificați de

fenotip, nu știm dacă părinții sunt heterozigoți și trebuie să genotipeze toți părinții din colecția noastră de triouri, chiar dacă unii se vor dovedi a nu fi informativi.

TDT a fost extins pentru a fi aplicabil în cazurile în care ADN-ul unuia sau ambilor părinți nu este disponibil, la frați și la alte grupuri de familie, precum și la trăsăturile cantitative. Un exemplu este programul FBAT (http://www.hsph.harvard.edu/fbat/default.html).

8.10 Studii de asociere la nivelul genomului: GWAS

În loc să investigăm o singură genă candidată, putem efectua studii de asociere la nivel de genom (GWAS) pentru a testa polimorfisme în întregul genom. [37] Primul lucru de remarcat este că un GWAS nu acoperă toate variațiile genetice ale genomului. În schimb, un GWAS examinează un set mare de SNP (de obicei de ordinul a 500.000 până la câteva milioane) distribuite în genom pentru a acoperi variația comună (alele care sunt purtate de 1% sau mai mult din populație). Aceste SNP-uri sunt genotipizate folosind micromatrice ADN (numite uneori „cipuri SNP"). Micromatricele mai noi includ, de asemenea, variante mai rare care apar în exom; exomul constă din 1–2 % din genom care codifică de fapt proteine.

Deși genomul conține multe milioane de SNP-uri, putem selecta un subset mai mic pentru a acoperi variația comună la nivelul genomului, deoarece alelele la multe SNP-uri dintr-o anumită regiune sunt corelate și, prin urmare, poartă informații redundante. Acest lucru se datorează fenomenului de dezechilibru de legătură pe care l-am menționat mai devreme. Dezechilibrul legăturii se referă la asocierea non-aleatorie (adică, corelația) dintre alele care sunt fizic apropiate pe un cromozom și sunt moștenite împreună. Cu alte cuvinte, din cauza dezechilibrului de legătură, un SNP poate reprezenta (sau „eticheta") un set mai mare de SNP. Ca rezultat, dacă aflăm că un SNP este asociat cu fenotipul nostru, acesta poate sau nu să fie legat cauzal de fenotip, ci mai degrabă poate fi pur și simplu corelat (adică, în dezechilibrul de legătură) cu adevăratul SNP cauzal.

În plus față de SNP-urile direct genotipizate pe o microarray, este acum o practică standard de a extinde acoperirea SNP-urilor negenotipate prin efectuarea imputării. Acest lucru este posibil prin eforturile - inclusiv Proiectul Internațional HapMap și Proiectul 1000 de genomi - care au genotipizat sau secvențializat genomurile indivizilor din populațiile din întreaga lume. Aceste mostre servesc ca panouri de referință care includ mult mai multe SNP decât cele incluse într-un anumit experiment. Cunoscând modelele de dezechilibru de legătură ale SNP-

urilor din aceste panouri de referință, putem imputa alele la SNP-uri pe care nu am fost genotipați direct, recuperând un număr mult mai mare de SNP-uri pentru experimentul nostru.

Analiza GWAS este de obicei efectuată folosind pachete statistice specializate, dintre care cel mai utilizat este cunoscut sub numele de PLINK (http://pngu. mgh.harvard.edu/~purcell/plin/).

Un avantaj major al GWAS și al altor metode la nivel de genom este că putem cerceta întregul genom pentru a identifica SNP-urile asociate fără a fi nevoie să pariezi în avans pe genele candidate. Ca rezultat, GWAS este adesea menționată ca o abordare „impărținată" a asocierii – adică nu se bazează pe ipoteze anterioare care se bazează de obicei pe cunoașterea incompletă a bazei biologice a unei boli sau trăsături.

Studii recente au arătat că cele mai comune SNP identificate de GWAS nu fac parte din partea de codificare a proteinelor a genelor. În schimb, se găsesc mai ales în regiunile ADN despre care se crede că sunt implicate în reglarea activității genelor (expresia genelor). După cum sa menționat mai sus, totuși, un SNP asociat poate fi doar în dezechilibru de legătură cu adevăratul SNP cauzal (sau poate fi un fals pozitiv așa cum vom discuta mai târziu). Identificarea adevăratei surse a unui semnal de asociere necesită astfel studii suplimentare. Acestea pot include studii de cartografiere fină în care regiunea asociată este genotipată sau secvențiată mai profund (vezi secțiunea 8.17) pentru a restrânge semnalul sau studiile biologice în care SNP-urile asociate sunt examinate pentru efectul lor funcțional posibil asupra expresiei genelor sau alte teste biologice.

8.11 Controlul calității GWAS și echilibrul Hardy–Weinberg

Înainte de a examina rezultatele unui GWAS, este crucial să curățați datele prin efectuarea unei serii de pași de control al calității. De obicei, începem prin a elimina SNP-urile și persoanele pentru care calitatea datelor este slabă. De exemplu, este obișnuit să se elimine SNP-urile pentru care lipsesc mai mult de 5% dintre genotipuri și apoi să se elimine indivizii pentru care lipsesc mai mult de 2% dintre tipurile geno . Apoi, putem elimina SNP-urile pentru care diferența dintre cazuri și controale în % genotipuri lipsă este mai mare de 2 %; de exemplu, ați elimina SNP-urile în care 10 % dintre genotipuri lipsesc în cazuri și doar 2 % lipsesc în controale. În plus, eliminăm SNP-urile care nu sunt în echilibrul Hardy-Weinberg. Echilibrul Hardy-Weinberg (HWE) se referă la ideea că, în ipoteza - împerecherii aleatorii, frecvențele genotipului rămân constante de-a lungul

generațiilor, cu excepția cazului în care unele tulburări evolutive perturbă echilibrul . Acest principiu este util pentru controlul calității genotipării, așa cum este descris mai jos.

Luați în considerare un locus SNP cu două alele, să spunem A și G. Frecvența lui A în populație este p și frecvența lui G este q. Deoarece la acel locus poți avea doar un A sau un G, atunci p + q = 1. Acum probabilitatea ca să obții o alele A de la mama ta și o alela A de la tatăl tău este produsul celor două probabilități, deoarece acestea sunt independente (a se vedea secțiunea 2.2). Deci probabilitatea ca genotipul tău să fie AA este

$$^{p}(\mathrm{O}_{\text{mamă}}) \times {}^{p}(\mathrm{Un}_{\text{tată}}) — {}^{p}:$$

Care este probabilitatea ca genotipul tău să aibă un A și unul G, adică să fii AG sau GA? Ei bine, există două moduri prin care poți avea o alela A și una G: obții un A de la mama ta și un G de la tatăl tău cu probabilitatea pq sau poți obține un G de la mama ta și un A de la tatăl tău tot cu probabilitate de p × q. Deci probabilitatea de a avea un A și unul G este 2pq.

Care este probabilitatea unui genotip GG? Aici trebuie să obțineți un G de la tatăl și mama dvs. și probabilitatea unei alele G este q, deci probabilitatea de GG este q × q sau q^2. Astfel, există trei genotipuri posibile :

AA, AG sau GG și suma totalului tuturor probabilităților posibile de genotip la acel loc este 1, deci $p^2 + 2pq + q^2 = 1$. Dacă vă amintiți algebra de bază, $p^2 + 2pq + q^2$ este expansiunea lui $(p + q)^2$.

Dacă proporțiile genotipului dintr-o probă sunt semnificativ diferite de $p^2 + 2pq + q^2$, trebuie să știm de ce. Este posibil ca genotipurile SNP să nu fie în echilibrul Hardy-Weinberg din cauza erorilor de genotipizare sau din cauza încălcării ipotezei de împerechere aleatorie. Astfel, trebuie să eliminăm aceste SNP-uri din setul nostru de date. Cu toate acestea, în studiile caz-control, un SNP care este cu adevărat asociat cu boala poate să nu urmeze HWE în proba combinată caz-control, deoarece alelele au distribuții diferite în cazuri și martori. Deci, testele HWE sunt de obicei efectuate pentru fiecare SNP din controale, iar cele care încalcă HWE sunt eliminate.

8.12 Quantile după Quantile Plots sau QQ Plots

Construirea unui diagramă QQ este adesea primul pas în evaluarea calității unui GWAS. Diagramele QQ ne permit să comparăm vizual două distribuții de statistici de test sau valori p. În GWAS, graficele QQ sunt adesea folosite pentru a examina dacă valorile p observate urmează distribuția așteptată în cadrul ipotezei nule generale. Abaterile de la distribuția nulă pot indica probleme cu datele, inclusiv inflația statisticilor de test din cauza stratificării populației , probleme cu imputarea, date rare sau erori de genotipizare.

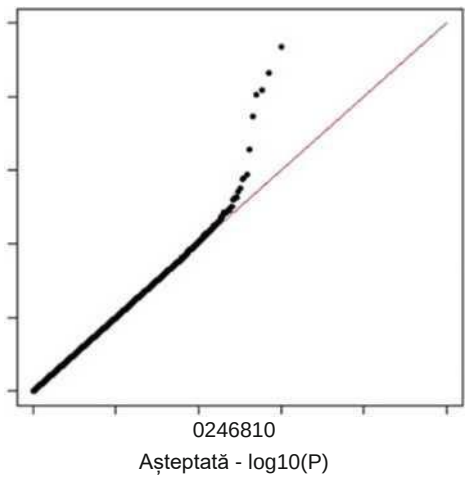
Să presupunem că faci un GWAS de 5.000 de cazuri și 5.000 de controale privind un milion de SNP. Fiecare dintre acele teste statistice SNP între cazuri și controale are valoarea ap asociată cu acesta. Valoarea p provine dintr-un test chi-pătrat care compară frecvența unei anumite alele în cazuri față de martori (sau poate proveni dintr-o regresie logistică, în care stabiliți raportul de șanse al bolii [starea cazului] asociată cu o anumită alele în timp ce controlând pentru alte variabile). În orice caz, puteți pune acele valori p pentru milionul de SNP-uri într-o listă ordonată, cu excepția faptului că, în loc să mergem de la cea mai mică valoare la cea mai mare valoare, o inversăm și ordonăm lista mergând de la cea mai mare valoare p la cea mai mică valoare p. valoare, de exemplu, trecând de la p = 1,0 la p = .000001.

O valoare p observată de 1,0 este la percentila 0 - ceea ce înseamnă că toate valorile p sunt mai mici decât p = 1,0 și nu există valori mai mari de p = 1,0 (în schema noastră, ele sunt la dreapta lui 1,0 în lista noastră ordonată - vezi exemplu de mai jos). Valoarea p de .000001 poate fi la percentila 99, ceea ce înseamnă că 99 % din valorile p sunt mai mari decât .000001 sau doar 1 % din valorile p sunt mai mici decât .000001 (în dreapta lui .000001). În realitate, le transformăm în – log p astfel încât p de 1,0 ar fi 0 și p de .000001 ar fi 6. Acest lucru este ilustrat în tabelul de mai jos.

P	1.00	0,10	0,05	0,01	0,001	0,0001	0,00001
Percentila	0	al 90-lea	al 95-lea	al 99-lea	al 999-	9999	99999
-log10 P	0	1	1.3	2	3	4	5

Acum luați în considerare distribuția așteptată a valorilor p dacă de fapt ipoteza nulă ar fi adevărată, ceea ce înseamnă că nu există o diferență generală

semnificativă între cazuri și martori, ci doar o variație aleatorie a milioanelor de valori p testului SNP din statisticile chi-pătrat. Acum reprezentați grafic distribuția observată a cuantilelor valorii p din GWAS-ul dvs. în raport cu distribuția teoretică așteptată, ca mai jos.



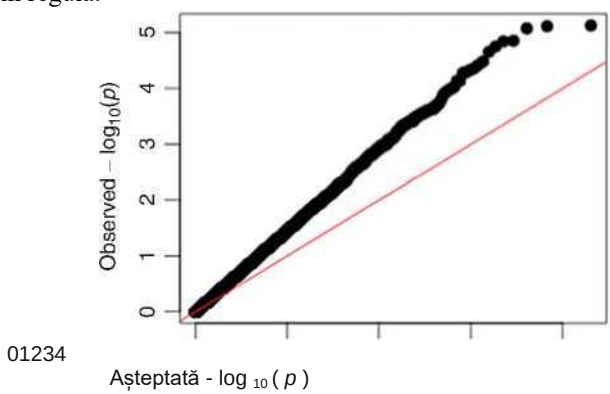
Dacă distribuția valorilor p în GWAS a fost exact ca distribuția așteptată a valorilor p în ipoteza nulă, punctele

s-ar afla de-a lungul liniei de referință roșie diagonală. Punctele în negru sunt ceea ce ați obținut de la GWAS. Ei urmează de-a lungul liniei roșii până în partea din dreapta sus, unde se află deasupra liniei roșii. Aceasta înseamnă că ați obținut mai multe puncte cu valori p foarte, foarte scăzute (sau valori mari de $-\log_{10} p$), decât v-ați aștepta să obțineți în ipoteza nulă. Deci, există o grămadă de teste SNP care au valori p mai mici decât cele așteptate și acestea pot reprezenta diferențe reale SNP între cazuri și controale.

Acum uitați-vă la un alt diagramă QQ de mai jos. Aici punctele din GWAS-ul dvs. se îndepărtează de linia de referință aproape de-a lungul celei mai mari a parcelei, ceea ce înseamnă că cuantilele din GWAS-ul dvs. nu corespund cuantilelor așteptate. Există mai multe teste SNP cu valoare p mai mare în GWAS decât v-ați aștepta într-o mare parte a valorilor p ordonate. Deci ceva nu este în regulă – probabil că există prea multe false pozitive , poate din cauza erorilor de laborator, a stratificării populației sau a altor motive, cum ar fi relația genetică

ascunsă între subiecții tăi. Aceasta nu este o situație fericită, deoarece pune la îndoială testele SNP semnificative pe care le-ați găsit. Ceea ce doriți să vedeți într-un diagramă QQ sunt puncte cât mai aproape posibil de linia diagonală în cea mai mare parte a distribuției, pentru a indica faptul că nu există erori sau părtiniri sau confuzie între cazuri și controale. Deci, diagramele QQ sunt un fel de metodă de control vizual al calității.

De asemenea, putem cuantifica inflația statisticilor testului folosind λ (vezi Secțiunea 8.8) În cazul unui studiu GWAS, λ este mediana chi-pătratului testelor între cazuri și controale pentru toți markerii, împărțit la 0,456, care este mediana distribuției chi-pătrat sub nul. Un λ de 1,0 indică lipsa inflației. Pentru graficul QQ prezentat mai sus, λ este 1,1 sugerând o inflație minimă. Cu toate acestea, pentru graficul QQ prezentat mai jos, λ este substanțial (1,788), ceea ce sugerează că ceva nu este în regulă.



După cum am menționat mai devreme (secțiunea 8.8), putem corecta pentru inflația statisticilor testului împărțind toate valorile chi-pătrat la λ . Cu toate acestea, o metodă mai puternică și mai comună de abordare a stratificării populației în GWAS implică controlul ascendenței genetice folosind multe mii de SNP care nu sunt asociate cu boala pe care o studiați. În esență, folosim aceste SNP-uri pentru a surprinde variația în ascendență și reducem această variație de fond genetic la un număr gestionabil de variabile (de obicei 4-20), prin aplicarea „analizei componentelor principale". Aceste componente principale ale variației genetice pot fi apoi adăugate ca covariabile la testele de asociere de regresie logistică sau regresie liniară ale fiecărui SNP. Când un GWAS include numărul adecvat de componente principale, controlăm efectiv stratificarea populației și λ nu mai este umflat.

8.13 Probleme de fals pozitive

Amintiți-vă că într-un studiu de asociere încercăm să vedem dacă o anumită alele, să zicem alela G a unui SNP A/G, apare mai frecvent în cazuri decât la martori. Putem folosi un test statistic standard, cum ar fi chi-pătrat, pentru a compara frecvențele de caz și de control. Dar amintiți-vă problema testării multiple, care duce la o eroare mare de tip I, adică o probabilitate mare de a declara o diferență între cazuri și controale atunci când există

într-adevăr nu este nicio diferență (secțiunea 3,24). Am observat că, de exemplu, dacă facem 5 comparații în două direcții într-un studiu, probabilitatea ca cel puțin una dintre aceste comparații să atingă nivelul de semnificație de 0,05 doar întâmplător este într-adevăr $1–.95^{5}$, sau .23, care ar fi nivelul nostru de semnificație efectivă. Pentru a obține un nivel de semnificație general adevărat de .05, folosim corecția Bonferroni unde împărțim .05 la numărul de comparații sau .05/ 5 = .01, astfel încât trebuie să ajungem la un nivel de semnificație de .01 (mai degrabă decât .05) pentru a concluziona că există o diferență semnificativă în oricare dintre comparațiile în două sensuri.

Ce corecție ar trebui să folosim pentru un GWAS? Se pare că un studiu tipic GWAS implică efectiv un milion de teste independente. Rețineți că, din cauza structurii de dezechilibru de legătură a genomului (adică, corelația dintre SNP-urile din apropiere), acest număr efectiv de teste este adevărat dacă micromatricea noastră ADN include 500.000 SNP-uri sau două milioane SNP-uri. Dacă am folosi pragul tradițional al valorii p de 0,05, fără a corecta testele multiple, ne-am aștepta ca 5% sau 50.000 de SNP să depășească acel prag doar întâmplător! Deci pentru a evita falsele pozitive trebuie să aplicăm corecția Bonferroni și să împărțim .05 la 1 milion ceea ce ne oferă un prag statistic de $5 \times I0^{8}$. Astfel, trebuie să obținem o valoare ap de < .00000005 pentru a declara semnificația statistică a oricărui SNP dat. Desigur, dacă testăm mai multe fenotipuri, trebuie să corectăm în continuare numărul de fenotipuri examinate pentru a menține rata de eroare de tip I de 5%. De asemenea, trebuie remarcat faptul că pot apărea fals pozitive din cauza erorilor de genotipizare sau chiar din cauza dimensiunii insuficiente a eșantionului. Astfel, descoperirile semnificative la nivelul genomului sunt cele mai de încredere atunci când sunt obținute în studii bine dezvoltate după un control riguros al calității setului de date. În cele din urmă, validitatea unei asocieri nu este stabilită până când nu a fost replicată în eșantioane independente.

8.14 Problema falselor negative

Dacă faceți un GWAS și nu găsiți niciun SNP care să fie semnificativ diferit în cazuri față de controale (cu $p < 5 \times 10^{8}$), poate fi din cauza lipsei de putere. Amintiți-vă că puterea este probabilitatea de a găsi un efect când

există într-adevăr unul și depinde atât de dimensiunea efectului predictorului (aici un SNP) cât și de dimensiunea eșantionului. Dimensiunile tipice ale efectului în GWAS oferă cote mici, în intervalul 1,01–1,3. Puterea de a detecta un astfel de efect depinde de mai mulți factori: frecvența alelei, dimensiunea eșantionului, prevalența bolii (pentru o boală binară), raportul caz/martor și modul de moștenire al alelei de risc (aditiv, recesiv sau dominant) . Pentru a da un exemplu, detectarea unei alele cu o frecvență de 5 % care poartă un raport de șanse de 1,3 pentru o boală cu prevalență de 2 % conform unui model aditiv, am avea nevoie de 10.350 de cazuri dacă presupunem un număr egal de martori.[III] Pentru o alelă cu o frecvență de 10%, am avea nevoie de 5540 de cazuri folosind aceleași ipoteze pentru mărimea efectului, prevalența bolii, raportul caz/martor și modul de moștenire. De asemenea, putem calcula cotele minime detectabile pentru o anumită dimensiune a eșantionului. De exemplu, având în vedere 10.000 de cazuri și 10.000 de controale și presupunând o frecvență alelelor de 10%, o prevalență a bolii de 2% și un model aditiv, am putea detecta un raport de șanse de aproximativ 1,22. După cum arată clar aceste exemple, sunt necesare dimensiuni foarte mari ale eșantionului pentru a detecta tipul de efecte pe care ne așteptăm să le vedem într-un GWAS. Acesta este motivul pentru care în prezent majoritatea studiilor genetice s-au bazat pe colaborări sau consorții ale mai multor cohorte diferite pentru a obține o putere suficientă.

Unii au sugerat că pragul de semnificație al lui $5 \times 10 \wedge 8$ este prea strict și în loc de procedura simplă Bonferroni, ei folosesc o statistică numită rata de descoperire falsă, care este numărul de ipoteze nule respinse în mod fals împărțit la numărul total. de respingeri. Există și alte nuanțe în încercarea de a evita falsele

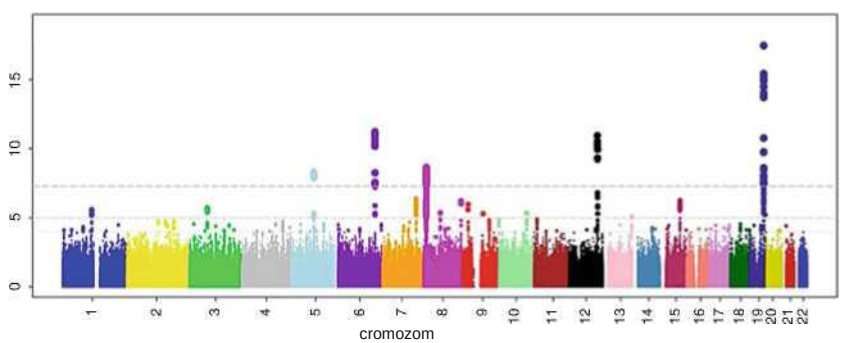
[III] Acest lucru presupune, de asemenea, un model multiplicativ pentru efect - adică, dacă purtarea unei alele de risc conferă un risc relativ de 1,3, purtarea a două alele de risc conferă un risc relativ de 1,69 (adică, 1,3 × 1,3).

pozitive, asigurând în același timp cea mai mare probabilitate de a găsi adevărate pozitive care depășesc scopul acestei cărți, dar referințele sunt furnizate în secțiunea de referințe. [38]

8.15 Manhattan Plots

Un diagramă Manhattan este un instrument de vizualizare care ne permite să rezumăm toate rezultatele asocierii unui GWAS într-un singur diagramă. Luați în considerare un studiu caz-control al geneticii accidentului vascular cerebral ischemic. Fiecare caz și fiecare control a fost genotipizat pe un cip microarray care conține un milion de SNP. Cum înțelegeți aceste un milion de teste? Graficul Manhattan este practic un grafic al valorilor p obținute de fiecare SNP pe care l-ați testat.

Pe axa X se află locația SNP-urilor de-a lungul genomului, ordonată în funcție de poziția de-a lungul cromozomilor. Imaginează-ți un milion de SNP-uri de pe cip întinse într-o linie lungă. Axa Y corespunde nivelului de semnificație pentru testul de asociere al fiecărui SNP; este — log al probabilității, astfel încât pentru valoarea ap de .0001, — log de .0001 = 4; pentru valoarea ap de .00001, — log de .00001 = 5; și așa mai departe. Deci, cu cât valoarea p este mai mică, cu atât este mai mare numărul –log(P)



Linia orizontală superioară corespunde valorii p pragului acceptat de 5×10^{-8}, care minimizează fals pozitive (a se vedea secțiunea 8.13). Orice lucru deasupra acestei linii este considerat suficient de puțin probabil să fi apărut întâmplător, astfel încât să putem presupune că există o diferență reală între cazuri și controale la acel SNP. În figura de mai sus, vedem cinci hit-uri semnificative la nivel de genom, adică regiuni care depășesc acest prag (câte unul pe cromozomii 6, 8, 12

și 19). Cel mai semnificativ SNP este cel mai înalt

punct, iar „coada” punctelor direct dedesubt corespunde SNP-urilor care sunt în dezechilibru de legătură cu SNP-ul superior.

8.16 Scoruri Polygene

Bolile sau trăsăturile complexe sunt de obicei considerate a fi poligenice - adică multe sute sau mii de loci pot contribui. Astfel , presupunem că multe SNP care nu reușesc să depășească semnificația la nivelul genomului (adică, $p < 5 \times 108$) sunt totuși asociate cu adevărat cu fenotipul de interes. Astfel, pe lângă descoperirea de noi asociații prin testarea fiecărui SNP, GWAS poate fi folosit și pentru a genera scoruri agregate care surprind efectele multor alele într-o singură variabilă. Aceste „scoruri poligene” pot fi apoi utilizate ca predictori sau covariabile în studiile viitoare ale unei boli în eșantioane independente care au fost GWAS ' d. Cu alte cuvinte, scorul poligenelor oferă o singură măsură a încărcării genetice pentru alelele de risc comune.

Iată cum să construiți un scor poligen . Începem cu un GWAS al bolii de interes, să spunem infarctul miocardic (IM), în eșantionul nostru primar sau „de descoperire”. Apoi clasăm SNP-urile după valorile lor p de la cel mai mic la cel mai mare. Folosim doar SNP ale căror valori p sunt mai mici decât un anumit prag, să spunem $p < .0001$; de fapt, de obicei construim scoruri pentru câteva praguri diferite. Odată ce avem toate SNP-urile pe care le vom folosi, formula pentru scor este

$$\text{Polygenscore} = \Sigma \left({}_{\text{ito}}\, j \right) x_i \log \left(\text{SAU}_i \right)$$

unde x este numărul de alele de risc (0, 1 sau 2) pe care un individ le poartă la al i -lea SNP; SAU_i este raportul de cote alelic pe care l-am găsit pentru acel SNP; și i la j înseamnă că însumăm toate SNP-urile sub pragul nostru de valoare p, până la j-lea SNP final.

Acum aplicăm formula scorului poligen pentru fiecare persoană din noul nostru set de date „țintă” independent pentru a obține scorul pentru toate SNP-urile. Adesea, ponderile scorului poligenelor pentru o boală au fost deja calculate din studii GWAS anterioare, iar ponderile ($\log(OR_i)$) pot fi obținute fără a fi

nevoie să facem noi înșine o descoperire GWAS.

Acum, imaginați-vă că efectuăm un studiu pentru a examina efectul unei expuneri (de exemplu, evenimente stresante de viață, LES) asupra riscului de infarct miocardic (IM). Ar fi obișnuit să controlăm factorii de risc cunoscuți în regresia noastră, astfel încât modelul nostru să fie scris ca

$$Y = \alpha + \beta_1 \text{LES} + \beta_2 \text{fumat} + \beta_3 \text{colesterol} + \beta_4 \text{diabet} + \beta_5 \text{hipertensiune arterială}$$

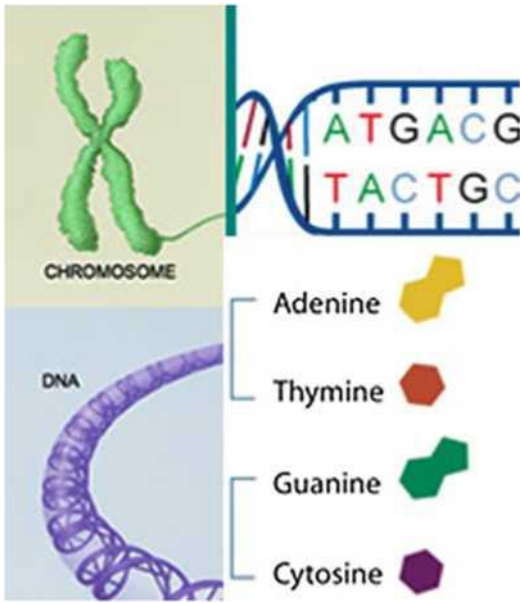
unde Y = caz vs. control și β ' s sunt coeficienți pentru fiecare predictor sau covariabilă. Acum putem adăuga scorul poligen pentru MI (derivat dintr-un GWAS anterior) ca covariabilă:

$$Y = \alpha + \beta_1 \text{LES} + \beta_2 \text{fumat} + \beta_3 \text{colesterol} + \beta_4 \text{diabet} + \beta_5 \text{hipertensiune} + \beta_6 \text{scor poligen}$$

În această regresie, putem estima efectul independent al scorului poligenului asupra depresiei și, de asemenea, putem estima efectul independent al LES asupra controlului riscului de depresie pentru factorii de risc, inclusiv încărcarea genetică (adică, scorul poligenului) pentru IM.

8.17 Variante rare și secvențierea genomului

După cum am spus, GWAS care utilizează micromatrice ADN sunt concepute pentru a testa SNP-uri comune în întregul genom. Cu toate acestea, ele nu acoperă toate variațiile care există în genom. Pentru a măsura toate variațiile, ar trebui să examinăm secvența completă a genomului uman. Secvențierea implică măsurarea secvenței complete a celor patru nucleotide A, G, C și T într-o regiune a ADN-ului; vezi figura de mai jos.



Prin amabilitatea Bibliotecii Naționale de Medicină și
a Institutului Național al Genomului Uman, NIH

Proiectul Genomului Uman (HGP) a secvențiat întregul genom uman de aproximativ trei miliarde de baze, derivate din ADN-ul mai multor voluntari anonimi. Această „secvență de referință" (rs) a fost finalizată în 2003. (Amintiți-vă că, odată ce cunoașteți secvența de nucleotide dintr-o catenă de ADN, o cunoașteți în catena de potrivire, deoarece A se împerechează întotdeauna cu T și G se împerechează întotdeauna cu C) . Mai recent, un efort la scară mai mare de a cataloga variante rare prin secvențierea ADN-ului a fost întreprins prin Proiectul 1000 de genomi. Acest proiect este o întreprindere internațională de cartografiere a variantelor din genomul uman care apar cu o frecvență de cel puțin 1 %, pentru ca cercetătorii să poată studia aceste variante în legătură cu boala. Acum este clar că bolile complexe implică atât variante comune (care individual au efecte mici și sunt capturate într-un GWAS), cât și variante mai rare (care individual pot avea efecte moderate până la

efecte mari, dar nu sunt măsurate într-un GWAS tipic). În iunie 2014, Proiectul 1000 Genomes a secvențializat genomurile a 2.535 de persoane din 26 de populații din întreaga lume, rezultând un catalog de 79 de milioane de variante de ADN.

8.18 Analiza studiilor de variante rare

Prima secvență a genomului uman a necesitat mai mult de un deceniu pentru a fi produsă la un cost de aproximativ 1 miliard de dolari . De atunci, progresele tehnologice au redus considerabil costurile și au crescut viteza de secvențiere. Costul secvențierii unui întreg genom a scăzut acum sub 1.000 de dolari . Studiile de secvențiere a ADN-ului iau de obicei una dintre cele două forme. Primul implică secvențierea tuturor exonilor (porțiunea care codifică proteinele din genom, cunoscută sub numele de „exom"). Deși exomul reprezintă mai puțin de 2 % din genom, variațiile exonilor sunt mai ușor de interpretat. Asta pentru că atunci când o mutație perturbă un exon, este ușor de văzut cum poate duce direct la o boală. A doua abordare implică secvențierea întregului genom (adică „secvențierea întregului genom"). Deși acest lucru oferă mult mai multe informații decât secvențierea întregului exom, poate fi mai greu să fii sigur că o schimbare în altă parte a genomului are un impact funcțional.

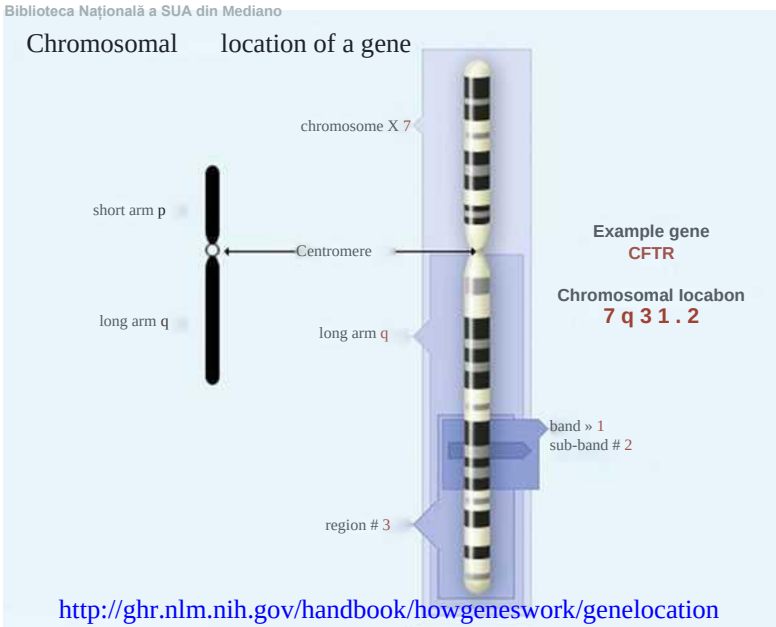
După cum am subliniat mai devreme, variantele rare detectate prin secvențiere pot avea efecte mult mai mari decât variantele comune care arată asociere într-un GWAS. Variantele cu efecte mari sunt de așteptat să fie rare, deoarece selecția naturală va fi împiedicat alelele dăunătoare să devină comune într-o populație. Deoarece efectele mai mari sunt mai ușor de detectat decât cele mai mici, ați putea crede că studiile de secvențiere nu ar avea nevoie de dimensiunile mari ale eșantioanelor necesare pentru GWAS. Și asta e adevărat … uneori. De exemplu, mutații rare în bolile mendeliane (o singură genă) au fost descoperite prin studierea câtorva familii care poartă mutația. Dar pentru boli complexe precum diabetul sau schizofrenia – unde pot fi implicate multe variante rare diferite, dintre care unele nu pot fi nici măcar moștenite (mutații de novo) – problema mărimii eșantionului este încă o provocare. Asta pentru că variantele rare sunt rare. Puterea de a găsi a
asocierea genetică depinde atât de mărimea efectului, cât și de frecvența unei variante. Deci, așa cum avem nevoie de multe mii de cazuri și controale pentru a

găsi o variantă care este comună, dar are un efect mic (GWAS), avem nevoie de numere similare (de ordinul a 25.000 de cazuri) pentru a găsi o variantă care este rară, chiar dacă are un efect mare.

Pentru variantele rare, testele standard de asociere pe care le folosim în GWAS nu vor funcționa pentru dimensiunile eșantionului care sunt disponibile în acest moment. O variantă cu o frecvență de 0,1% va apărea de numai 10 ori la 10.000 de indivizi. Deci, o strategie comună este de a prăbuși variante rare într-o genă (sau cale biologică) și de a analiza datele la nivelul genelor. De exemplu, ne putem întreba dacă „povara" variantelor rare este mai mare în rândul cazurilor decât în cazul controalelor. Variantele pot fi, de asemenea, ponderate prin probabilitatea lor de a fi importante din punct de vedere funcțional sau dăunătoare (de exemplu, dacă ar perturba o proteină). Putem apoi corecta pentru teste multiple folosind o corecție Bonferroni pentru numărul de gene din genom, adică 0,05/20.000.

8.19 Ce este într-un nume? SNP-uri și gene

HUGO este Organizația Genomului Uman. Comitetul de nomenclatură a genelor HUGO (HGNC) este responsabil pentru furnizarea de linii directoare pentru denumirea genelor umane și pentru aprobarea numelor noi, unice ale genelor umane și a abrevierilor de formă scurtă. Figura de mai jos ilustrează semnificația numelui: 7q31.2. Este locația citogenetică pe cromozomul genei CFTR, care este gena pentru fibroza chistică. Simbolurile 7q31.2 înseamnă că se află pe al șaptelea cromozom, în brațul lung al cromozomului (q), în banda 1, sub-banda 2.



Genomul uman secvențiat de Proiectul Genomul Uman este reprezentat de o singură secvență cunoscută sub numele de secvență de referință , care este un compus de date de la un număr de voluntari anonimi. Un SNP, de exemplu, poate avea numele rs2532087. „rs" înseamnă „SNP de referință" și numărul este atribuit de grupul NCBI dbSNP (National Center for Biotechnology Information database of Single Nucleotide Polymorphisms). Aceste numere sunt pur și simplu identificatori (nu poziții genomice).

A existat o schimbare de paradigmă în știință – de la a crede că lucrurile sunt mai simple decât par la a înțelege că sunt mai complexe decât par. În ultimul secol, principiul care a călăuzit efortul științific a fost briciul lui Occam - că explicația cea mai parcimonioasă pentru fenomene este cea mai bună. Dar, pe măsură ce descoperirile genomice și moleculare se accelerează, devine evident că, în sfera biologică, explicațiile simple nu sunt posibile și scopul este de a descoperi și explica mai precis complexitatea (și minunea) inerentă a vieții.

Chapter 9 PREDICȚIE DE RISC SI CLASIFICAREA RISCURILOR

> Este mult mai bine să prevedeți chiar și fără certitudine decât să nu prevedeți deloc. –Henri Poincare în Fundamentele științei, p. 129.

9.1 Predicția riscului

Suntem interesați de prezicerea riscului unei boli pentru o persoană, deoarece deciziile de tratament se bazează adesea pe risc. Vom folosi riscul cardiovascular ca exemplu în secțiunile următoare, deoarece predicția riscului este cea mai dezvoltată pentru această boală. Ghidurile de tratament pentru hipertensiune arterială sau colesterol ridicat de la Asociația Americană a Inimii diferă pentru persoanele cu risc crescut de boli cardiovasculare de cele cu risc mai scăzut. De exemplu, medicamentele anticoagulante sunt recomandate persoanelor care au fibrilație atrială (un tip de aritmie cardiacă) dacă prezintă un risc crescut de accident vascular cerebral, determinat de vârsta lor, dacă au diabet, antecedente de accident vascular cerebral, insuficiență cardiacă și hipertensiune arterială. Deoarece anticoagulantele prezintă un risc de sângerare, acestea nu sunt recomandate persoanelor care au un risc scăzut de accident vascular cerebral. Predicția riscului este, de asemenea, foarte utilă în probleme de sănătate publică. A ști ce procent din populație este expusă unui risc ridicat poate ajuta planificatorii sănătății să adopte măsuri preventive și să planifice utilizarea resurselor.

Factorii de risc acceptați în mod tradițional pentru accident vascular cerebral, în general, disponibili pentru clinicieni, sunt vârsta, tensiunea arterială sistolică, diabetul zaharat, fumatul, bolile cardiovasculare anterioare, fibrilația atrială, hipertrofia ventriculară stângă prin electrocardiogramă și utilizarea medicamentelor antihipertensive . Ecuațiile de estimare a riscului de accident vascular cerebral au fost dezvoltate din studii prospective precum Studiul Framingham (a se vedea secțiunea 4.10) în care o populație specificată din orașul Framingham Massachusetts a avut măsurători ale acestor variabile și au fost urmărite în timp pentru a vedea cine a dezvoltat evenimente precum accidentul vascular cerebral. sau atac de cord și cât de bine variabilele de bază studiate ar

putea prezice cine va dezvolta evenimentul sau rezultat. [39-41] Odată cu apariția unor tehnici mai sofisticate de măsurare a anumitor proteine din sânge (biomarkeri), oamenii au devenit interesați de perfecționarea modelelor de predicție a riscului și de a vedea dacă anumiți biomarkeri ar putea îmbunătăți predicția. Întrebarea este cum evaluăm dacă sau cât de mult noul biomarker îmbunătățește predicția riscului? Este o întrebare importantă, deoarece, dacă biomarkerul ajută la predicție, acesta poate fi utilizat în mod obișnuit în cabinetele medicilor . Adăugarea de biomarkeri la predicția riscului poate indica faptul că unii oameni, despre care se credea anterior a fi cu risc scăzut, sunt expuși unui risc mai mare și ar trebui tratați. Pe de altă parte, acest lucru se poate adăuga la costurile de îngrijire a sănătății, așa că dacă nu îmbunătățește predicția și nu are nicio influență asupra deciziilor de tratament, este posibil să nu merite să faceți testul în mod obișnuit.

Prima indicație că un biomarker poate fi util este că este asociat în mod semnificativ cu rezultatul de interes, fie într-o regresie logistică , fie într-un model de regresie a hazardelor proporționale Cox (a se vedea Secțiunea 4.19 pentru explicația modelelor de hazarduri proporționale Cox). Punctul important de remarcat este că un biomarker poate fi asociat în mod semnificativ cu un rezultat al bolii, dar asta nu înseamnă că îmbunătățește neapărat predicția riscului sau reclasifică oamenii în diferite categorii de risc.

Ca exemplu, să luăm în considerare o proteină numită CRP (proteina C-reactivă), care este un marker al inflamației generale și poate fi analizată dintr-un simplu test de sânge. Vrem să vedem dacă se adaugă la predicția accidentului vascular cerebral.

Situația ideală ar fi dacă am măsura biomarkerul de interes pentru toată lumea înainte de eveniment și apoi i-am urmărit înainte în timp, deoarece unii au dezvoltat rezultatul, iar alții nu (cu alte cuvinte, un studiu prospectiv) și apoi am adăugat biomarkerul la variabile. în modelul nostru de predicție a riscului. Dar devine foarte costisitor să faci asta, așa că abordarea obișnuită este să faci un studiu caz-control și să măsori biomarkerul numai în cazuri și în controalele acestora. Acest capitol discută măsurile pe care le folosim pentru a evalua valoarea aditivă a unui biomarker în predicția riscului.

9.2 Valoarea aditivă a unui biomarker: calcul a riscului prezis

Vom utiliza în continuare predicția riscului de accident vascular cerebral ischemic ca exemplu, utilizând date dintr-un studiu caz-control al accidentului vascular cerebral ischemic imbricat în cadrul studiului prospectiv de observare a inițiativei pentru sănătatea femeilor asupra femeilor aflate în postmenopauză . [42] Considerăm biomarkerul CRP (proteina C-reactivă), un marker al inflamației. Au existat 868 de cazuri și 883 de controale cărora li s-a testat CRP la examinarea inițială și care nu au avut antecedente de accident vascular cerebral. (De fapt, cele 883 de cazuri de accident vascular cerebral au fost corelate cu 883 de controale pe mai multe variabile, dar 15 dintre cazuri nu au avut o probă de sânge adecvată, rămânând 868 de cazuri și 883 de controale).

Mai întâi trebuie să dezvoltăm o ecuație de predicție a riscului din datele noastre și apoi trebuie să estimăm probabilitatea de accident vascular cerebral într-o anumită perioadă de timp pentru fiecare persoană din populația noastră de studiu folosind ecuația de predicție a riscului fără biomarkerul din ea. Această probabilitate va depinde de valorile specifice pentru fiecare persoană ale variabilelor pe care le folosim în modelul de predicție. Ideea generală este că mai întâi trebuie să clasificăm oamenii în categorii de risc pe baza modelului nostru vechi fără CRP. În continuare, clasificăm aceleași persoane pe baza noului nostru model, care constă din variabilele din modelul vechi plus biomarkerul CRP. Comparăm cele două scheme de clasificare cu rezultatele reale și vedem ce model prezice mai precis. Secțiunile de mai jos descriu mai multe măsuri utilizate pentru a evalua cât de bine se adaugă biomarkerul la predicție. Câteva lucrări excelente ale lui Pencina intră în mai multe detalii. [43–45]

Mai întâi rulăm un model de regresie logistică din care putem calcula probabilitatea de accident vascular cerebral. Folosim variabile utilizate în mod tradițional în predicția accidentului vascular cerebral în studiul Framingham: vârsta, tensiunea arterială sistolică , diabetul zaharat, fumatul, bolile cardiovasculare anterioare, fibrilația atrială, hipertrofia ventriculară stângă prin electrocardiogramă și utilizarea medicamentelor antihipertensive într-o ecuație. pentru a prezice apariția viitoare a accidentului vascular cerebral. (În exemplul nostru am omis HVS pentru că nu aveam date în acest sens și pentru că prevalența HVS în populația WHI a fost foarte scăzută. În plus, utilizarea medicamentelor antihipertensive este o variabilă care depinde de nivelul tensiunii arteriale, dar pentru

simplificare considerăm doar medicația = 1 dacă femeia o folosește și = 0 dacă nu).

Amintiți-vă din capitolul 4 că într-o regresie logistică, probabilitatea unui eveniment este

$$Prevent) \quad \frac{1}{1 + e^{-k}}$$

$$unde\ k = C_0 + C_1 X_1 + C_2 X_2 + C_3 X_3 + \ldots + C_m X_m$$

Parametrii în k (valorile lui C_0 la C_m) sunt obținute din regresia logistică care poate fi rulată în diverse pachete statistice computerizate precum SAS sau STATA. În studiul nostru caz-control de accident vascular cerebral, regresia logistică necondiționată pe care am calculat-o cu variabilele utilizate în scorul de risc Framingham, *k* , a fost

k = (-2,4421 - .0108 (vârstă) + .0207(TAS) + .0003 (la medicație pentru tensiunea arterială) + .2228 (antecedente de CHD) + .9868 (fumător actual) + .5829 (fibrilație atrială) + . 7867 (diabet) + .2411 (caucazian))

Dar trebuie să facem o ajustare la *k* pentru a ține seama de faptul că este un studiu caz-control. Astfel, adăugăm termenul *ln* α- p : n-Δ-nrols-) la interceptarea modelului de regresie logistică necondiționată; P este probabilitatea de accident vascular cerebral la populația țintă. Într-un studiu prospectiv, putem obține P direct, deoarece va fi doar numărul de cazuri de accident vascular cerebral în perioada de urmărire (alegem 8 ani). Dar deoarece acesta este un studiu caz-control și P prin definiție este de 50%, adică există 50% cazuri de accident vascular cerebral și 50% controale în proiectul nostru, trebuie să estimăm P pentru populația țintă. Pentru WHI estimăm P ca incidența anuală a accidentului vascular cerebral în WHI

Studiu observațional, care a fost de 0,0029 anual, ori mai mare decât media de urmărire de 8 ani sau 0,0232. Astfel, factorul nostru de corecție este

$$\text{correction factor ln} \quad \frac{P}{1 \quad P} \quad \frac{n_{controls}}{: n_{cases}} = ln(\nearrow 0232\lambda / 883 \; :9768 : 868$$

= ln(: 02375 *ori* 1 : 01728) = ln (: 02416) = -3 : 7230

Rețineți dacă există un număr egal de cazuri și controale, atunci $\frac{ncontrols}{n\ cazuri}$ devine 1 și factorul de corecție este pur și simplu logaritmul natural al lui 1 - P

K corectat corectat = k + factor de corecție

Apoi determinăm probabilitatea de accident vascular cerebral în decurs de 8 ani pentru fiecare persoană din studiul nostru, utilizând valorile variabilelor respectivei persoane . Deci, de exemplu, pentru al i -lea persoană care este o femeie albă, nefumătoare în vârstă de 55 de ani, cu tensiune arterială sistolică de 120, care nu are diabet zaharat, nu ia medicamente antihipertensive, fără fibrilație atrială sau antecedente de boli de inimă, se calculează probabilitatea de accident vascular cerebral în 8 ani după cum urmează:

$$k_{\text{corectat}} = (-2{,}4421 - .0108\,(55_{\text{de ani}}) + .0207(120_{\text{sbp}}) + .0003\,(O_{\text{medicamente}}) + .2228\,(0_{\text{chd}}) + .9868\,(0_{\text{fumător curent}}) + .5829\,(0_{\text{fibrilația atrială}}) + .7867\,(0_{\text{diabet zaharat}}) + .2411\,(1_{\text{Caucazian}}) + (-3{,}7230_{\text{co rr ectionfactor}}) = -4{,}0340$$

$$\text{și } P_i = 1 - e^{-k} = 1 - 56.4862^{57.4862 \cdot 01}\,7^{1}$$

Deci persoana respectivă are un risc estimat de accident vascular cerebral în 8 ani de 1,7%.

Calculăm aceste probabilități pentru fiecare dintre cazurile de accident vascular cerebral și fiecare dintre controale. Apoi împărțim aceste probabilități în categorii de risc. Am ales următoarele categorii de risc: < 2 %, 2 % până la < 5 %, 5 % până la < 8 % și ≥8 %. Categoriile acceptate de obicei sunt < 5 % risc scăzut, 5 până la mai puțin de 10 % risc mediu scăzut, 10-20 % risc intermediar ridicat și > 20 % risc ridicat. Dar, în exemplul nostru, avem o populație cu risc în general scăzut până la mediu (în virtutea faptului că i-am exclus pe toți cei cu un accident vascular cerebral sau un atac de cord anterior). Categoriile de risc pe care le-am ales corespund aproximativ nivelurilor de risc scăzute, intermediare și ridicate utilizate în deciziile de inițiere a tratamentului pentru prevenirea accidentului vascular cerebral la persoanele cu fibrilație atrială. Persoana din exemplul nostru de mai sus se află în categoria cu cel mai scăzut risc de < 2 %.

Repetăm același proces folosind noul model care include variabila CRP. Rețineți că acest nou model cu CRP are coeficienți ușor diferiți pentru fiecare dintre variabile decât modelul vechi fără CRP. (Folosim lnCRP — jurnalul natural al CRP, deoarece CRP nu este distribuit în mod normal).

Să presupunem că femeia noastră din eșantion are un nivel CRP de 12 (care este foarte mare):

$$\ln(12.0) = 2.4849$$

În exemplul nostru, $k_{cRp,\ corectat}$ = (-2,7246 - .0078(55 $_{a\,g\,e}$) + .0201(120 $_{sb\,p}$) + .0000(0$_{medications}$) + .2048(0$_{history\ of\ chd}$) + .9789(0$_{current\ smoker}$) + .5615 (0 $_{fibrilație\ atrială}$) +.7003(0 $_{diabet\ zaharat}$) +.2076(1 $_{caucazian}$)) +.1954 (2.4849 $_{lnCRP}$) + (—3,7230) = —3,7714 și probabilitatea ei estimată de accident vascular cerebral în 8 ani este prezentată mai jos:

$$p^{i(CRP)} \quad \frac{1}{1+e^{-k}} \quad \frac{1}{1+44.4746} \quad \frac{1}{45.4746} \quad '0215$$

sau o probabilitate de 2,2% de accident vascular cerebral în 8 ani.

Adăugarea nivelului CRP la model a crescut riscul ei scăzut de la 1,7 % la 2,2 %. Astfel, ea a trecut de la categoria cu cel mai scăzut risc (< 2 %) la o categorie cu risc mai mare (2 % până la < 5 %). Acest lucru îl poate determina sau nu pe medicul ei să-și schimbe tratamentul, dar îl poate determina pe medicul să facă o supraveghere mai activă a factorilor ei de risc. Făcând aceleași calcule pentru o femeie de 65 de ani, caucazian, diabetică, cu o tensiune arterială sistolică de 170 al cărei risc cu modelul vechi era de 8,2 % și al cărei CRP era de 5,0 (încă mai mare decât 3,0 considerat în intervalul normal, dar nu la fel de mare ca CRP-ul de 12 din exemplul nostru anterior), nu schimbă cu adevărat riscul ei, care este acum 8,3 %. Deoarece ea avea un risc mare de la început, deși CRP-ul ei a fost mai mare decât normal, dar nu îngrozitor, ea rămâne cu risc crescut, iar ceilalți factori de risc, cum ar fi diabetul și hipertensiunea arterială, copleșesc efectele CRP crescute. Acum, desigur, ceea ce vrem să știm este că aceste femei sunt cazuri sau controale? Trebuie să ne testăm previziunile în raport cu rezultatele reale.

După calcularea riscului cu și fără CRP pentru toate femeile din studiu, în continuare încrucișăm categoriile de risc pentru cele două modele separat pentru cele care au avut un accident vascular cerebral în perioada de 8 ani și pentru cele care nu au suferit un accident vascular cerebral. tabelul de mai jos:

Model vechi fără CRP Risc previzionat cu noul model care include CRP

Cazuri de accident vascular cerebral

Categorii de risc	<2%	2 până la <5%	5 până la <8%	>=8%	total
<2%	**6**	3	0	0	9
2 până la	20	**336**	67	0	423
5 până la	0	44	**157**	42	243
>=8%	0	0	20	**173**	193
total	26	383	244	215	868

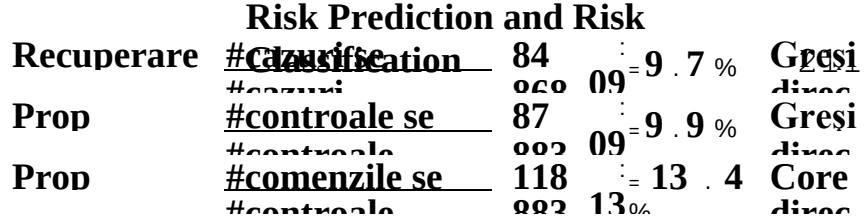
Controale

Categorii de risc	<2%	2 până la <5%	5 până la <8%	>=8%	total
<2%	**23**	10	0	0	33
2 până la	63	**480**	51	0	594
5 până la	0	44	**107**	26	177
>=8%	0	0	11	**68**	79
total	86	534	169	94	883

Rețineți că numerele de pe diagonală reprezintă femei al căror risc estimat nu a fost modificat prin adăugarea CRP la model. De exemplu, au existat 336 de cazuri de accident vascular cerebral cu risc estimat de 2 % până la < 5 % atât cu modelul vechi, cât și cu cel nou și 480 dintre controalele fără accident vascular cerebral care au rămas în aceeași categorie de risc. Numerele de deasupra diagonalei reprezintă persoane al căror risc a crescut odată cu adăugarea biomarkerului la model, iar cele de sub diagonală reprezintă persoane al căror risc prezis a scăzut cu noul model care include biomarkerul.

Ceea ce am dori să vedem cu un nou biomarker este că cazurile de accident vascular cerebral ar trebui să crească într-un risc prezis atunci când biomarkerul este adăugat și controalele fără accident vascular cerebral ar trebui să scadă într-un risc prezis. Cu alte cuvinte, dorim ca riscul prezis să se miște în direcția corectă: cazurile de accident vascular cerebral ar trebui să se miște în sus, iar controalele ar trebui să se miște în jos. Din exemplul nostru avem următoarele: (proporție (în sus|caz) înseamnă proporția care crește riscul între cazuri; proporție (în jos|control) înseamnă proporția care se mișcă în jos în risc între controale).

Recuperare #cazuri se 112 = 12 . 9 Core
#cazuri 868 12% direc


Risk Prediction and Risk
Recuperare #cazuri se Classification 84 = 9 . 7 % Gresi
#cazuri 868 09
Prop #controale se 87 = 9 . 9 % Gresi
#controale 883 09
Prop #comenzile se 118 = 13 . 4 Core
#controale 883 13%

9.3 Indicele net de îmbunătățire a reclasificării

NRI este = (prop (uplcase) — prop (downlcase)) — (prop (uplcontrol) — prop (jos|control))

În exemplul nostru este (.129 — .097) — (099 — .134) = .032 — (—.035) = .032 + .035 = .067 = 6.7 %

Aceasta înseamnă că, prin adăugarea biomarkerului, am reclasificat corect 6,7 % dintre oameni (adică, am avut o îmbunătățire netă a clasificării). O altă modalitate de a scrie acest lucru este mai jos, unde *p* este probabilitatea estimată (pe care o estimăm din eșantionul nostru analizând proporția), D = D este boală (sau accident vascular cerebral) = 1 și D = 0 fără boală (sau fără accident vascular cerebral) = 0:

$$NRI = [p(sus|D=1) - p(jos|D=1)] + [p(jos|D=0) - p(sus|D=0)]$$

Dar este acest lucru semnificativ diferit de zero? Pentru a afla, trebuie să calculăm scorul az (vezi capitolul 3) care este NRI împărțit la eroarea sa standard. Formula este dată mai jos.

$$z = \frac{NRI}{\sqrt{\frac{p(sus|D=1)+p(jos|D=1)}{\#\ evenimente} + \frac{p(sus|D=0)+p(jos|D=0)}{\#\ non\text{-}evenimente}}}$$

În exemplul nostru,

$$z = \frac{.129 + .097, .099 + .134}{\sqrt{\frac{.129 + .097}{868} + \frac{.099 + .134}{883}}} = \frac{.067}{\sqrt{.000260 + .000264}} = \frac{.067}{\sqrt{.000524}} = \frac{.067}{.02289}$$

$$= 2.93$$

Z de 2,93 are valoarea ap de < .01; astfel, NRI este semnificativ diferit de 0.

NRI este o măsură a valorii aditive a unui biomarker la predicția riscului. Un dezavantaj, însă, este că depinde de categoriile de risc absolut pe care le selectăm. NRI fără categorii este un alt indice, care nu depinde de modul în care categorizăm riscul.

9.4 NRI fără categorie

NRI-ul fără categorii se uită doar la mișcarea riscului prezis în sus sau în jos atunci când biomarkerul este în model, indiferent dacă această mișcare înseamnă trecerea persoanelor la o altă categorie de risc. Este procentul tuturor subiecților ale căror estimări de risc sunt modificate în direcția corectă minus procentul schimbat în direcția incorectă pentru a obține o cifră a efectelor nete. Direcția corectă pentru cazuri este riscul crescut în modelul cu biomarker comparativ cu modelul fără biomarker, iar pentru controale este riscul scăzut în modelul cu biomarker comparativ cu modelul fără biomarker. Deci, pentru fiecare persoană trebuie să calculăm probabilitatea de accident vascular cerebral folosind valorile ei ale variabilelor din modelul cu și fără biomarker și apoi numărăm.

Se calculează ca

$$\frac{\text{numărul de cazuri al căror risc cu biomarker este } \mathbf{mai\ mare} \text{ decât riscul fără biomarker}}{\text{numărul total de cazuri}}$$

$$\frac{\text{numărul de controale al căror risc cu biomarker este } \mathbf{mai\ mic} \text{ decât riscul fără biomarker}}{\text{numărul total de controale}}$$

$$\frac{\text{numărul de cazuri al căror risc cu biomarker este } \mathbf{mai\ mic} \text{ decât riscul fără biomarker}}{\text{numărul total de cazuri}}$$

$$\frac{\text{numărul de controale al căror risc cu biomarker este mai mare decât riscul fără biomarker}}{\text{numărul total de controale}}$$

Practic, vedem ce proporție de oameni și-a schimbat riscul în direcția corectă (în creștere pentru cazuri, în jos pentru controale) și scădem proporția de oameni care și-a schimbat riscul în direcția greșită (în scădere pentru cazuri și în sus pentru controale).

În exemplul nostru, îmbunătățirea netă a reclasificării fără categorii a fost de 18,9 %, ceea ce înseamnă că estimarea riscului pentru 18,9 % dintre oameni a fost modificată cu orice sumă în direcția corectă atunci când CRP a fost adăugat la modelul de predicție.

9.5 Îmbunătățirea integrată a discriminării (IDI)

O calitate pe care o dorim în adăugarea unui nou biomarker la o ecuație de predicție este că ar trebui să îmbunătățească discriminarea dintre cei care suferă un accident vascular cerebral de la cei care nu . Un indice al discriminării este măsura de îmbunătățire a discriminării integrate (IDI), în care comparăm probabilitățile medii estimate pentru cazuri și controale. Pentru a calcula IDI, mai întâi calculăm două cantități pentru fiecare persoană: (1) probabilitatea de accident vascular cerebral pentru fiecare persoană utilizând modelul de regresie logistică fără CRP în el (model vechi) și (2) probabilitatea de accident vascular cerebral folosind modelul cu CRP în el. (model nou). Obținem apoi media acestor probabilități pentru cazurile de accident vascular cerebral și media pentru controale. IDI este diferența dintre probabilitățile medii prezise între cazuri și controale folosind noul model minus diferența dintre probabilitățile medii prezise între cazuri și controale folosind modelul vechi.

Notăm p = probabilitatea estimată

$$\text{IDI} \quad \text{IJmCan}^{p}_{(\text{model}_{\text{nou}}, \text{c}_{\text{as}}\text{e}_{\text{s}}} \quad \text{mean } p\,(\text{∏ew model, controls}) \quad \text{^^^^^B}$$

$$\text{m}^{\text{ea}}\text{∏ } p\,(\text{old model, controls})$$

$$^{\text{mea}}\text{∏}^{p}_{(\text{old modei , cases})}$$

IDI pentru exemplul nostru = $(0:0688_{\text{model nou , cazuri}} - 0:0474_{\text{model nou , controale}})$ $(0:0661_{\text{model vechi , cazuri}} - 0:0475_{\text{model vechi , controale}}) = 0:0028$ Putem testa ipoteza nulă că IDI = 0 prin calcularea Z (vezi capitolul 3) ca

$$Z = \frac{\text{IDI}}{/\,SE^{2}(\text{cazuri}) + SE^{2}(\text{controale})}$$

Pentru a obține SE (cazuri) obținem diferențele pentru fiecare persoană dintre

probabilități pentru modelul vechi și modelul nou. Apoi primim

eroarea standard a acestor diferențe. (Reamintim că $SE = \frac{stand}{ar\ d}$, abatere). (√ n)
SE (controale) este calculată în același mod, dar folosind diferențele pereche de probabilități în controale.

În exemplul nostru, IDI de .0028 este semnificativ statistic cu $p < .001$.

9.6 C-Statistică

O altă măsură a discriminării este c-statistica, care este probabilitatea ca o persoană selectată aleatoriu cu evenimentul de interes (accident vascular cerebral în exemplul nostru) să aibă un risc prezis mai mare decât o persoană aleasă aleatoriu fără eveniment. Statistica c reflectă aria de sub curbă (AUC) dintr-o curbă caracteristică de operare a receptorului (ROC) (vezi secțiunea 5.2). (Curba ROC este un grafic al sensibilității pe axa y vs. 1-- specificitatea pe axa x). Valorile c-statistice mai mari indică o mai bună discriminare . În general, statistica c nu se mișcă foarte mult cu adăugarea unui biomarker și este oarecum greu de interpretat în acest context. Cu toate acestea, este utilizat și raportat pe scară largă, chiar dacă poate să nu fie o măsură foarte bună a valorii predictive adăugate a unui nou biomarker. SAS va tipări statistica c atunci când rulează o analiză de regresie logistică, la fel ca STATA.

9.7 Avertismente

Rețineți că ecuația de predicție pe care o obțineți din propriile date va prezice întotdeauna mai bine decât atunci când o aplicați unei alte populații. Așa că ar trebui să-l validezi pe un alt eșantion. Sau, dacă setul de date este suficient de mare, puteți face validarea internă dezvoltând modelul pe jumătate din eșantion și testându-l pe cealaltă jumătate.

Rețineți, de asemenea, că un nou biomarker este util din punct de vedere clinic dacă are ca rezultat o schimbare a tratamentului sau o altă acțiune, cum ar fi o supraveghere crescută. Dacă nicio decizie de tratament nu va fi afectată de îmbunătățirea predicției riscului,

atunci ați putea la fel de bine să utilizați vechile variabile pe care să vă bazați

deciziile de tratament, deși noul biomarker ar putea fi încă interesant din punct de vedere științific.

9.8 Rezumat

Pentru a evalua capacitatea de predicție adăugată a unui biomarker nou, luăm următorii pași:

1. Verificați dacă biomarkerul este asociat în mod semnificativ cu rezultatul de interes (din literatură sau din propriul studiu).
2. Dezvoltați un model de predicție a riscului folosind variabile acceptate în mod tradițional, dar fără biomarker. Utilizați modele de regresie a riscurilor proporționale Cox dacă aveți date prospective. Utilizați modele de regresie logistică cu un factor de corecție dacă aveți un studiu caz-control.
3. Vedeți dacă statistica c pentru noul model cu biomarker crește semnificativ față de modelul vechi fără biomarker (secțiunea 9.6).
4. Calculați probabilitățile de accident vascular cerebral pentru fiecare persoană din modelul cu și fără biomarker.
5. Calculați NRI (vezi Secțiunea 9.3).
6. Calculați NRI fără categorie (secțiunea 9.4).
7. Calculați IDI (secțiunea 9,5).
8. Raportați toate aceste măsuri deoarece nicio măsură nu este perfectă.

Chapter 10
ETICA ȘI STATISTICA CERCETĂRII

Moralitatea, ca și arta, înseamnă să tragem o linie undeva.
Oscar Wilde (1854–1900)

10.1 Ce legătură are statisticile cu ea?

La prima vedere poate părea că statistica și etica cercetării nu au nimic de-a face una cu cealaltă. Nu așa! Luați în considerare de ce atât de mulți oameni se oferă voluntari pentru studii de cercetare medicală. În multe cazuri, se datorează faptului că există un beneficiu așteptat. De exemplu, în studiile clinice pentru cancer, deseori, medicamentul de investigație este o ultimă speranță și este posibil să nu fie disponibil în afara studiului. În multe studii privind bolile cardiovasculare, participanții apreciază îngrijirea și atenția suplimentară și sunt dispuși să încerce un medicament nou, de exemplu, pentru hipertensiune arterială. Și, de fapt, s-a demonstrat că adesea, participanții la studiile clinice trăiesc mai mult și se descurcă mai bine decât populația generală, chiar dacă sunt tratați cu un placebo. Dar ceea ce poate nu este suficient de apreciat este faptul că mulți, mulți oameni participă la studii din altruism pentru a avansa cunoștințele științifice. Cunoștințele științifice nu sunt avansate atunci când un studiu este prost conceput sau realizat fără suficientă rigoare sau nu este suficient de mare pentru a oferi un răspuns. Statisticile adecvate sunt un factor determinant al eticii unui studiu.

Un prim exemplu este Inițiativa pentru sănătatea femeilor (WHI), descrisă în capitolul 6 . Femeile aflate în postmenopauză au fost rugate să se alăture unui studiu de terapie de substituție hormonală; studiul va continua până la 12 ani înainte ca rezultatele să fie cunoscute și s-ar putea să nu beneficieze în mod direct femeile, dar ar răspunde la întrebarea importantă a efectului hormonilor asupra cancerului, bolilor de inimă și osteoporozei. Mulți dintre participanții WHI au participat pentru fiicele și nepoatele lor și și-au exprimat mândria și entuziasmul pentru a răspunde la întrebările pentru generațiile viitoare. Și într-adevăr au atins acest obiectiv – o parte din WHI, studiul estrogen plus progestativ versus placebo, a răspuns deja la aceste întrebări importante, cu un rezultat uluitor: estrogen plus

 progestinul crește riscul de cancer de sân și, de asemenea, crește atacurile de cord, accidentul vascular cerebral și cheaguri de sânge și demența. Deci, deși tratamentul arată beneficii în ceea ce privește cancerul colorectal și fracturile osteoporotice, riscurile generale depășesc beneficiile. Acest proces a schimbat practica medicală pentru generațiile viitoare.

Ei bine, asta ne readuce la statistici. Acest studiu a putut să răspundă la aceste întrebări deoarece a avut suficientă putere pentru a le răspunde. A fost nevoie de 16.608 de femei din acea parte a WHI pentru a putea detecta aceste efecte. Chiar dacă rezultatul ar fi fost nul (adică, dacă nu a arătat nicio diferență între grupul de tratament și placebo), am fi putut avea încredere în acel rezultat, deoarece puterea era acolo de a detecta un efect adevărat dacă exista într-adevăr unul. După cum sa dovedit, rezultatele au fost clare, deși neașteptate, în favoarea placebo. Deci, ideea este că, pentru ca un studiu să fie „etic", trebuie să fie proiectat și alimentat suficient de bine pentru a putea răspunde la întrebările pe care le pune. În caz contrar, persoanele care consimt să participe la așteptarea că vor contribui la cunoaștere ar putea să nu contribuie de fapt, deoarece studiul este prost conceput, alimentat sau executat și pot fi expuși inutil la riscuri.

Rețineți că există anumite proiecte de studiu pentru care considerentele de putere sunt mai puțin relevante. Exemple sunt studiile pilot, care, prin definiție, sunt menite să testeze fezabilitatea unui protocol de cercetare sau să culeagă date preliminare pentru a planifica un studiu complet și sunt scutite de problema de putere. Considerațiile de putere pot să nu se aplice, de asemenea, anumitor studii de toxicitate a medicamentelor (studii de fază I) sau anumitor tipuri de studii de cancer, dar cu siguranță, în studiile de prevenire, precum și în studiile de tratament de fază III, puterea este o considerație majoră în etica cercetării.

10.2 Protecția subiectelor de cercetare umane

Subiecții umani din cercetarea medicală contribuie foarte mult la îmbunătățirea sănătății oamenilor. Acești voluntari trebuie să fie protejați de vătămări cât mai mult posibil. În trecutul nu prea îndepărtat, au existat unele încălcări flagrante ale principiilor etice în efectuarea cercetării medicale. Cele mai îngrozitoare exemple din lume sunt experimentele medicale efectuate

în lagărele de concentrare naziste — de medici! Sfidează orice fel de înțelegere a modului în care profesioniștii educați, probabil „civilizați”, ar fi putut să-și denatureze profesia și propria umanitate. Dar aceste atrocități au avut loc și au demonstrat ororile pe care oamenii sunt capabili să le comită. Când aceste atrocități au devenit cunoscute după cel de-al Doilea Război Mondial, procesele de la Nuremberg ale criminalilor de război naziști (inclusiv medicii care au efectuat astfel de cercetări) au dus și la Codul de la Nürnberg pentru conducerea cercetării medicale, care a stabilit cerința de bază a consimțământului informat voluntar. Ulterior, Declarația de la Helsinki, din 1964, a extins și rafinat liniile directoare de cercetare și a devenit un standard mondial, care suferă revizuiri periodice.

Cel mai infam exemplu de cercetare lipsită de etică din Statele Unite a fost probabil studiul Tuskegee Institute despre sifilis, care a avut loc în sudul Statelor Unite între 1932 și 1972. Cercetătorii au vrut să studieze cursul natural al sifilisului. În anii 1940 au devenit disponibile antibiotice care ar putea trata această boală, dar au fost refuzate participanților, care erau bărbați săraci de culoare, deoarece o intervenție pentru tratarea bolii ar interfera cu acest studiu observațional. În 1972 publicul a luat cunoștință de acest experiment și în 1974 a fost înființată Comisia Națională pentru Protecția Subiecților Umani de Cercetare Biomedicală și Comportamentală. Ei au elaborat un raport cunoscut sub numele de Raportul Belmont: principii etice și linii directoare pentru protecția subiecților umani de cercetare. Aceste linii directoare sunt urmate de toate școlile de medicină și alte instituții de cercetare care desfășoară cercetări care implică participanți umani și sunt considerate a fi principii universale, depășind liniile culturale. Orientările se bazează pe trei principii de bază: respectul pentru persoane, binefacerea și justiția.

Respectul pentru persoane recunoaște că oamenii sunt ființe autonome și pot face propriile lor alegeri informate cu privire la participarea la cercetare, fără constrângere. Procesul de consimțământ informat se bazează pe acest principiu. Participanții care nu sunt capabili să facă propriile alegeri, cum ar fi pacienții în comat sau persoanele cu incapacități mintale sau copiii mici , trebuie să aibă protecții speciale.

Beneficiul, sau principiul neafectiunii, înseamnă că riscurile cercetării trebuie să fie reduse la minimum, beneficiile maximizate, iar cercetătorul este responsabil pentru protejarea participantului.

Justiția în acest context se referă la o distribuție echitabilă a riscurilor și beneficiilor cercetării. Un grup de oameni nu ar trebui să fie expus riscurilor de

cercetare în beneficiul altui grup de oameni. Acesta poate deveni un concept destul de complicat. Deși poate fi ușor să discerneți încălcări în anumite situații - pentru a lua exemplul cel mai extrem, prizonierii naziștilor au fost supuși experimentelor de înghețare în beneficiul soldaților care ar putea fi nevoiți să lupte în condiții arctice - poate fi mai subtil în multe situații și acestea. trebuie examinat cu atenție, conform acestui principiu.

10.3 Consimțământ informat

Unul dintre cele mai importante elemente în protecția subiecților umani este principiul consimțământului informat. Subiectul studiului trebuie să își dea liber consimțământul să facă parte din studiu după ce a fost pe deplin informat cu privire la potențialele riscuri și beneficii.

Există anumite elemente care trebuie să fie într-o formă scrisă de consimțământ. Trebuie precizat scopul cercetării; trebuie listată o persoană de contact de 24 de ore; trebuie să existe o descriere a procedurilor de studiu: ce se așteaptă de la participant, durata studiului și cât timp va dura participantului . Riscurile și disconforturile potențiale, beneficiile potențiale și inconvenientele pentru participanți trebuie să fie toate clar precizate. Trebuie să existe o declarație conform căreia participarea este voluntară și că participantul are dreptul de a se retrage în orice moment și că acest lucru nu va prejudicia grija participantului. Dacă cercetarea poate duce la nevoie de îngrijire suplimentară sau proceduri de diagnosticare, participantului trebuie să i se spună în ce măsură este responsabil pentru îngrijirea ulterioară și pentru ce va plăti studiul. Dacă există vreo compensație pentru participanți, fie pentru cheltuielile efectuate în timpul participării, fie pentru timpul petrecut, aceștia trebuie să fie informați cu privire la sumă. (Suma nu trebuie să fie excesivă, deoarece poate părea coercitivă .) Trebuie inclusă o declarație care să asigure confidențialitatea și modul în care va fi păstrată.

Cel mai important, participantul trebuie să înțeleagă cu ce este de acord și formularul de consimțământ trebuie să fie redactat într-un limbaj care să fie înțeles și, dacă este cazul, tradus în limba maternă a participantului.

limbă. Toate acestea trebuie să fie aprobate de IRB (sau Comitetul de revizuire instituțional) al instituției medicale, care este, în general, un comitet de experți și persoane neprofesionale care analizează și trebuie să aprobe toate protocolurile de

cercetare înainte de începerea cercetării și care monitorizează evenimentele adverse ca cercetarea progresează. Diferitele IRB au cerințe specifice diferite , care sunt de obicei postate pe site-urile lor web. Consimțământul informat este un proces continuu – nu este doar semnarea unui formular la începutul unui studiu. Cercetatorul are obligația de a informa participantul cu privire la noile cercetări relevante care i-ar putea afecta decizia de a continua să participe.

Revenind la WHI, din moment ce se credea la momentul în care a fost început WHI că hormonii ar proteja femeile de bolile de inimă, formularul inițial de consimțământ a declarat acest lucru ca un beneficiu potențial. Riscurile potențiale menționate în formularul de consimțământ au inclus o creștere a cancerului de sân și a cheagurilor de sânge. Când WHI era în desfășurare, HERS (Heart and Estrogen Replacement Study) a publicat rezultate care indică faptul că pentru femeile care aveau deja boli de inimă (trial de prevenire secundară), înlocuirea hormonală nu a oferit niciun beneficiu. Ei au observat mai multe atacuri de cord în partea incipientă a studiului, cu o posibilă reducere târzie, ducând la nicio diferență generală între grupurile de tratament și placebo până la sfârșitul studiului. Aceste informații au fost furnizate printr-o corespondență specială tuturor femeilor care participă la programul hormonal WHI pentru prevenirea primară a bolilor de inimă. (Prevenția primară înseamnă că studiul a fost efectuat pe femei în general sănătoase.) Ulterior, datele timpurii din WHI însuși au indicat că a existat prejudiciu precoce în ceea ce privește bolile de inimă. Din nou, femeile au fost informate printr-o corespondență specială, telefon și discuții personale cu personalul clinicii. În cele din urmă, studiul estrogen plus progestativ a fost oprit după 5,2 ani (în loc de media planificată inițial de 8,5 ani), deoarece riscul de cancer de sân în exces a depășit o limită de oprire predeterminată și un indice global al efectelor generale a sugerat mai mult rău decât beneficiu, iar toate femeile în studiu au fost întrerupte de la pastilele lor de studiu.

10.4 Echilibru

Asta ne aduce la un alt concept: când este etic să începem un studiu clinic al unui nou tratament? Când există echilibru. Echilibrul înseamnă că există dovezi aproximativ egale că tratamentul poate oferi beneficii și că nu va oferi beneficii. Dacă suntem siguri că tratamentul oferă beneficii, nu ar trebui să-l refuzăm persoanelor care ar primi placebo în cadrul studiului. Bineînțeles că putem greși. Au existat critici la adresa Women 's Health Initiative care au spus că nu este etică să se facă un astfel de studiu, deoarece era bine cunoscut că hormonii protejează împotriva bolilor de inimă și ar fi lipsit de etică să refuzi acești hormoni femeilor

randomizate la placebo! Bineînțeles că acum știm că a fost greșit - grupul placebo s-a descurcat mai bine. La momentul începerii WHI, dovezile observaționale indicau beneficii în ceea ce privește bolile de inimă, dar nu fusese niciodată testată într-un studiu clinic, care este „standardul de aur". Astfel, au fost mulți oameni care nu au crezut că beneficiile înlocuirii hormonale erau deja stabilite de studiile observaționale și se dovedește că au dreptate. Cercetătorul, a cărui obligație este să protejeze participanții umani la cercetare , trebuie să creadă că este la fel de probabil că tratamentul este mai bun sau că placebo sau tratamentul de comparație este mai bun. Comunitatea științifică care judecă propunerea de cercetare trebuie să creadă, pe baza „statului actual al tehnicii", că există o întrebare rezonabilă la care trebuie să se răspundă.

10.5 Integritatea cercetării

Pentru ca concluziile cercetării să fie valide, procedurile de colectare a datelor trebuie să fie administrate riguros și uniform. Nicio dată nu poate fi modificată fără documente. Dacă există o eroare materială, modificarea și motivul acesteia trebuie documentate. Înscrierea trebuie să fie conform standardelor stricte și planificate în prealabil. Uneori (din fericire, rar) există o presiune mare pentru înscrierea subiecților într-un anumit interval de timp, sau cercetătorul (cu încălcarea principiului echilibrului) crede cu adevărat că tratamentul își poate ajuta pacienții și astfel „întorcă" regulile de înscriere. . Acest lucru poate invalida cercetarea și, prin urmare, nu este etică. Un exemplu foarte trist în acest sens a avut loc în cadrul Proiectului Național Chirurgical Adjuvant Breast and Bowel (NSABP). Acest studiu multicentric a demonstrat că lumpectomia ar putea fi echivalentă cu mastectomia la sute de mii de femei. Președintele acestui studiu a descoperit că investigatorul principal dintr-unul dintre centrele clinice a falsificat unele dosare ale pacienților, astfel încât femeile care nu erau eligibile pentru a participa la studiu pe baza unor criterii de înscriere predeterminate au fost făcute fals eligibile pentru a participa. Acest studiu excelent și extrem de important a fost inițial afectat când acest lucru a devenit cunoscut, iar președintele studiului a fost acuzat de Oficiul de Integritate a Cercetării (ORI) pentru abatere științifică, chiar dacă a notificat NIH cu privire la problemă când a aflat despre aceasta. Ulterior, a fost eliberat complet și i s-au oferit mai multe scuze. Studiul a avut implicații profunde asupra tratamentului femeilor cu cancer de sân. Cu toate acestea, aceasta

a fost o încălcare gravă a eticii din partea unui anchetator într-unul dintre numeroasele centre care ar fi putut invalida constatările. Din fericire, rezultatele au durat chiar și atunci când au fost excluși toți pacienții din clinica infracțională.

10.6 Politici de autor

În cercetarea medicală, majoritatea articolelor de cercetare originale au mai mulți autori, deoarece cercetarea medicală este un efort de colaborare. Majoritatea revistelor medicale și a instituțiilor de cercetare au politici specifice și stricte de autor (publicate în reviste și/sau pe site-uri web), dintre care multe includ următoarele elemente: (1) coautorii trebuie să aducă o contribuție intelectuală la lucrare (de exemplu, să conceapă cercetarea). , să efectueze analize, să scrie secțiuni ale lucrării sau să aducă contribuții editoriale), (2) toți coautorii trebuie să poarte responsabilitatea pentru conținutul acesteia și (3) coautorii trebuie să dezvăluie potențiale conflicte de interese (de exemplu, relevante sprijin din industrie, lectorate, deținere de acțiuni). Ordinea autorului poate fi uneori un punct de disputa și ar trebui discutată de coautori la începutul procesului.

10.7 Comitete de monitorizare a datelor și siguranței

În general, studiile clinice au un comitet de monitorizare a datelor și siguranței (DSMB) pentru a supraveghea studiul. Acestea sunt grupuri independente de experți în disciplinele relevante, care au o calitate consultativă. Sarcina lor este să monitorizeze procesul și să asigure siguranța participanților. Într-un studiu în orb, ei sunt singurii care văd datele neoarbe la intervale regulate, prestabilite. Dacă găsesc beneficii sau vătămări excesive într-un braț al studiului, ei ar sfătui să oprească studiul (așa cum sa întâmplat în Inițiativa pentru sănătatea femeilor) . De obicei, criteriile pentru oprirea unui studiu din cauza vătămării în grupul de tratament sunt mai stricte decât oprirea pentru beneficii.

10.8 Rezumat

Conducerea etică a cercetării are multe componente. Apar întrebări etice noi și dificile pe măsură ce știința avansează și noile tehnologii devin disponibile. Acest scurt capitol începe să vă ofere o idee despre unele dintre problemele implicate.

Informații mult mai detaliate sunt disponibile de pe diverse site-uri web, iar NIH are un curs online în protecția subiecților umani. IRB-urile locale vă pot oferi informații și surse suplimentare.

Post-scriptum
CÂTEVA COMENTARII DE DESPĂZIRE ASUPRA IMPACTULUI EPIDEMIOLOGIEI ASUPRA VIEȚILOR OMULUI

În urmă cu zece ani, o femeie cu cancer de sân ar fi probabil să facă o mastectomie radicală, care, pe lângă îndepărtarea sânului și desfigurarea rezultată, ar include și îndepărtarea unei mari părți a peretelui muscular din piept și ar lăsa-o incapacitată în multe feluri. Astăzi, aproape nimeni nu face o mastectomie radicală și mulți nici măcar nu fac o mastectomie modificată, dar, în funcție de cancer, poate face o lumpectomie care doar îndepărtează nodul, lăsând sânul intact. Cu ani în urmă, nimeni nu a acordat prea multă atenție radonului, un gaz inert eliberat din sol și disipat prin crăpăturile fundației în case. Acum este recunoscut ca principala cauză a cancerului pulmonar. Rolul nutriției în prevenirea bolilor nu a fost recunoscut de comunitatea științifică. De fapt, oamenii care credeau în importanța nutrienților în cauza și vindecarea bolii erau considerați a fi niște înfățișați, doar puțin nebuni. Acum este adesea subiectul articolelor, cărților și știrilor, iar sume substanțiale de fonduri de cercetare sunt investite în studii nutriționale. Astfel de studii influențează legislația, cum ar fi, de exemplu, reglementările conform cărora alimentele procesate trebuie să aibă etichetare standard, ușor de înțeles de publicul larg, a conținutului de grăsimi din alimente, precum și de sodiu, vitamine și alți nutrienți. Toate acestea au un impact asupra schimbării obiceiurilor alimentare ale populației, precum și asupra economiei industriei alimentare.

În domeniul sănătății, schimbări în tratament, prevenire și cunoștințe predominante apar atunci când există o confluență de circumstanțe: se dobândesc noi informații pentru a înlocui teoriile existente; există o diseminare a acestor informații către comunitatea științifică și către publicul larg; și există climatul psihologic, economic și politic adecvat care ar saluta adoptarea noilor abordări. Miologia epidemiilor joacă un rol major prin furnizarea metodelor prin care se dobândesc noile cunoștințe științifice. Adesea, primele indicii asupra cauzalității vin cu mult înainte ca un mecanism biologic să fie cunoscut. În jurul anului 1850, la Londra, dr. John Snow, consternat de suferința și decesele cauzate de epidemiile de holeră, a studiat cu atenție rapoartele unor astfel de epidemii și a remarcat că

holera era mult mai probabil să apară în anumite părți ale Londrei decât în alte

pătrat în care holera era răspândită și unde era mai puțin, și a remarcat că casele furnizate cu apă de o companie, Southwark and Vauxhall Company, aveau mult mai multe cazuri de holeră decât cele furnizate de o altă companie. De asemenea, știa că compania Vauxhall folosea ca sursă o zonă puternic contaminată de canalizare. Snow a insistat ca orașul să închidă pompa care furnizează apa contaminată, cunoscută sub numele de Broad Street Pump. Au făcut așa și holera s-a diminuat. Toate acestea au fost cu 25 de ani înainte ca cineva să izoleze bacilul holeric și cu mult înainte ca oamenii să accepte ideea că boala poate fi răspândită prin apă. În vremurile moderne, epidemia de SIDA este una în care metoda de răspândire a fost identificată înainte ca agentul infecțios, virusul HIV, să fie cunoscut.

Tehnicile epidemiologice au fost aplicate din ce în ce mai mult bolilor cronice, care diferă de bolile infecțioase prin faptul că pot persista o perioadă lungă de timp (în timp ce infecțiile de obicei fie ucid rapid, fie se vindecă rapid) și au, de asemenea, de obicei cauze multiple, dintre care multe sunt dificil de rezolvat. identifica. Și aici, epidemiologia joacă un rol central în identificarea factorilor de risc, cum ar fi fumatul pentru cancerul pulmonar. Astfel de cunoștințe sunt transpuse în acțiuni publice înainte ca toate căile biologice să fie elucidate . Acțiunea ia forma unor campanii educaționale, legi împotriva fumatului, restricții privind reclamele și alte mecanisme de limitare a fumatului. Factorii de risc pentru bolile de inimă au fost identificați prin studii epidemiologice clasice care au dus la schimbări ale stilului de viață pentru indivizi, precum și consecințe de politică publică.

Bolile cronice prezintă probleme diferite și provocatoare în analiză , iar noi tehnici statistice continuă să fie dezvoltate pentru a aborda astfel de probleme. De asemenea, sunt dezvoltate noi tehnici statistice pentru problemele speciale întâlnite în cercetarea genetică. Astfel, domeniul statisticii nu este static, iar domeniul epidemiologiei nu este fix. Ambele se adaptează și se extind pentru a face față problemelor de sănătate în schimbare ale societății noastre și cu progresele în cunoștințe, ca de exemplu în genetică, care necesită metodologii noi.

Appendix 1
VALORI CRITICE ALE CHI-PĂTRATUL, Z ȘI t

Când valoarea Z, χ^2 sau t calculată din datele observate este egală cu sau depășește valoarea critică enumerată mai jos, putem respinge ipoteza nulă la nivelul de semnificație dat, α (alfa).

Valori critice selectate ale chi-pătratului

Nivel de semnificație	.1	.05	.01	.001
Valoarea critică a lui χ^2	2,71	3,84	6,63	10.83

Valorile critice alese ale lui Z

Nivel de semnificație				
Test cu două cozi	.1	.05	.01	.001
(test cu o singură coadă)	(.05)	(.025)	(.005)	(.0005)
Valoarea critică a lui Z	1,64	1,96	2,58	3.29

Valorile critice alese ale t

Nivel de semnificație				
Test cu două cozi	.10	.05	.01	.001
(test cu o singură coadă)	(.05)	(.025)	(.005)	(.0005)
Grade de libertate				
9	1,83	2.26	3.25	4,78
19	1,73	2.09	3,86	3,88
100	1,66	1,98	2,63	3.39
1.000	1,64	1,96	2,58	3.29

NOTĂ: Interpretare:
Dacă ai 19 grade de libertate, pentru a respinge H_o, la $\alpha = .05$ cu un test cu două cozi, ai avea nevoie de o valoare a t mai mare sau mai mare de 2,09; pentru $\alpha = .01$, ar fi nevoie de cel puțin la fel de mare ca 3,86. Rețineți că atunci când df devine foarte mare, valorile critice ale lui t sunt aceleași cu valorile critice ale lui Z. Alte valori decât cele calculate aici apar în majoritatea textelor prezentate în lecturile sugerate.

Appendix 2
TESTUL EXACT AL FISHER

Să presupunem că aveți un tabel 2 pe 2 care rezultă dintr-un experiment pe șobolani care expune un grup la o anumită condiție experimentală și celălalt grup la o condiție de control, cu rezultatul de a fi în viață după o săptămână. Tabelul arată după cum urmează:

Tabelul A.1 Expunerea în funcție de rezultat

	Controla	Experimental	
În viaţă	*o* 1	*b* 7	Rândul 1 = $R_1 = 8$
Mort	*c* 5	*d* 1	Rândul 2 = $R_2 = 6$
Total	Col 1 = $C_1 = 6$	Col 2 = $C_2 = 8$	$N = 14$

87,5 % din grupul experimental și 16,7 % din grupul de control au trăit. Un rezultat mai extrem, având în vedere aceleași totaluri de rând și coloană, ar fi

Tabelul A.2 Expunerea în funcție de rezultat

	Controla	Experimental	
Alive	0	8	8
Dead	6	0	6
Total	6	8	14

unde au trăit 100 % din grupul experimental și 0 % din grupul de control. Un alt rezultat mai extrem ar fi locul în care au trăit 25 % din experimentul și toți martorii:

Tabelul A.3 Expunerea în funcție de rezultat

	Controla	Experimental	
În viaţă	0	2	8
Mort	0	6	6
Total	6	8	14

(Orice alte tabele pe care le-am putea construi cu aceleași totaluri marginale ar fi mai puțin extreme decât Tabelul A.1 , deoarece nicio celulă nu ar conține un număr mai mic decât cel mai mic număr din Tabelul A.1 , care este 1).

Calculăm probabilitatea exactă de a obține rezultatul observat al experimentului doar întâmplător (Tabelul A.1) sau una chiar mai extremă (ca în Tabelul A.2). sau A.3), dacă ar fi adevărat că nu au existat diferențe de supraviețuire între cele două grupuri. Testul exact al lui Fisher se calculează obținând probabilitatea fiecăruia dintre aceste tabele și însumând aceste probabilități.

Mai întâi trebuie să explicăm simbolul „!”. Se numește „factorial”. Un număr **n** ! înseamnă *(n)* × (**n** -1) × (**n** -2) × … × (1). De exemplu, 6! = 6 × 5 × 4 × 3 × 2 × 1. Prin definiție, 0! este egal cu 1.

Probabilitatea de a obține observațiile din tabelul A.1 este

$$\frac{(\mathbf{R}_1!)\times(\mathbf{R}_2!)\times(\mathbf{C}_1!)\times(\mathbf{C}_2!)}{\mathbf{a}!\times\mathbf{b}!\times\mathbf{c}!\times\mathbf{d}!\times\mathbf{N}!} = \frac{8!6!6!8!}{1!7!5!1!14!} = :015984$$

Probabilitatea de a obține observațiile din Tabelul A.2 este

$$\frac{8!6!6!8!}{0!8!6!0!14!} = :000333$$

Probabilitatea de a obține observațiile din tabelul A.3 este

$$\frac{8!6!6!8!}{6!2!0!6!14!} = :009324$$

Suma acestor probabilități este .015984+ .000333 + .009324 = .025641. Astfel putem spune că probabilitatea exactă de a obține rezultatele pe care le-am observat în Tabelul A.1 , sau rezultate mai extreme, este .025641, dacă ipoteza nulă a fost adevărată. Putem respinge ipoteza nulă că rata de supraviețuire este aceeași în ambele grupuri la un nivel de semnificație p = .026.

Appendix 3
TEST NEPARAMETRIC KRUSKAL-WALLIS PENTRU COMPARAREA MAI MULTE GRUPE

De exemplu, să presupunem că aveți trei grupuri de oameni fiecare având un scor pe o anumită scară. Numărul total de persoane din toate cele trei grupe este N. Procedura generală este următoarea: (1) Combinați toate scorurile din cele trei grupe și ordonați-le de la cel mai mic la cel mai mare. (2) Se acordă rangul de 1 celui mai mic punctaj, 2 următorului cel mai mic punctaj și așa mai departe, **N** fiind atribuit persoanei cu cel mai mare punctaj. (3) Sortați oamenii înapoi în grupurile lor originale, fiecare persoană având rangul său atribuit. (4) Însumați toate rangurile din fiecare grupă. (5) Calculați cantitatea prezentată mai jos, pe care o numim **H** . (6) Dacă aveți mai mult de cinci cazuri în fiecare grup, puteți căuta **H** într-un tabel chi-pătrat, cu **k** –1 grade de libertate (unde **k** este numărul de grupuri comparate).

Scoruri la înțelegerea citirului					
Grupa A		Grupa B		Grupa C	
Scoruri	(Rang)	Scoruri	(Rang)	Scoruri	(Rang)
98	(13)	80	(9)	120	(21)
70	(6)	60	(2)	110	(17)
68	(5)	106	(15)	90	(12)
107	(16)	50	(1)	114	(19)
115	(20)	75	(8)	105	(14)
65	(4)	74	(7)	85	(10)
Suma	**(64)**	64	(3)	112	(18)
rangurilor			**(45)**	87	(11)

					(122)

$$H = \left(\frac{12}{N(N+1)} \times \sum \frac{(R)^2}{n_j}\right) - 3(N+1)$$

$$H = \left(\frac{12}{21(22)}\right)\left(\frac{67.5^2}{6} + \frac{122^2}{8}\right) - 3(22)$$

grade de libertate = 3 — 1 = 2

Dacă ipoteza nulă a lipsei diferențelor de rang mediu între grupuri a fost adevărată, probabilitatea de a obține un chi-pătrat mai mare sau mai mare de 7,57 cu 2 grade de libertate este mai mică de .05, deci putem respinge ipoteza nulă. și concluzionăm că grupurile diferă. (Când apar egalități în clasament, fiecărui scor i se dă media rangului pentru care este egal. Dacă există multe egalități, se poate folosi o corecție la **H** , așa cum este descris în cartea lui Siegel enumerată în Lecturi sugerate).

Appendix 4
CUM SE CALCULEAZĂ UN COEFICIENT DE CORELARE

Individual	*X*	*Y*	X^2	Y^2	*XY*
O	5	7	25	49	35
B	8	4	64	16	32
C	15	8	225	64	120
D	20	10	400	100	200
E	25	14	625	196	350
S	73	43	1339	425	737

$$r \_ \quad N\Sigma XY - (\Sigma \mathbf{X})(\Sigma \mathbf{Y})$$

$$\Lambda / N\Sigma X^2 - (\Sigma X)^2 \Lambda / N\Sigma Y^2 - (\Sigma Y)^2$$

$$\_\_\_\_\_ \frac{5(737) - (73)(43)}{} \qquad \frac{3685 - 3139}{}$$

$$yj\ 5(1339) - (73)2\text{^}5(425) - (43)^2 \qquad \sqrt{T366}\sqrt{}^2 76$$

$$\frac{546}{(37)(16:6)} \quad \frac{546}{614} \_ : 89$$

Cum se calculează coeficienții de regresie

$$b \_ \quad \Sigma XY$$

1339 -ðʃ (73) 2 _ 737 - 628 _ 109 _ 4 0 ^ 1339
- 1066 ,, 273 ,, :

43 :40(73)
T 5 = 8 : 60 - 5 : 84 = 2 : 76

Appendix 5
AJUSTARE-VÂRSTA

Luați în considerare două populații, A și B, cu următoarele caracteristici:

Populația	Vârstă	Tarife specifice vârstei	# de persoane din populație	# de decese în populație	Rata brută a mortalit
O	Tineri	$1{,}0\,^{4}00\,^{=.004}$	500	.004 × 500 = 2	
	Vechi	$_{1.000}=.016$	500	.016 × 500 = 8	
	Total		1.000	10	10 / 1.000
B	Tineri	$1{,}0\,^{5}00\,^{=.005}$	667	.005 × 667 = 3.335	
	Vechi	$_{1,\,^{2}0\,^{0}00}=.020$	333	.020 × 333 = 6,66	
	Total		1.000	10	10 / 1.000

Rețineți că populația B are rate de deces specifice vârstei mai mari în fiecare grup de vârstă decât populația A, dar ambele populații au aceeași rată brută de deces de 10/1.000. Motivul pentru aceasta este că populația A are o proporție mai mare de bătrâni (50 %) și chiar dacă rata mortalității la bătrâni este de 16/1.000 în populația A față de 20/1.000 în populația B, numărul de persoane mai mare este în acel grup contribuie la un număr mai mare de decese totale.

Pentru a efectua ajustarea vârstei, trebuie să selectăm o populație standard cu care vom compara atât A cât și B. Următoarele exemple folosesc două populații standard diferite ca ilustrații. În practică, o populație standard este aleasă fie ca populație într-un anumit an, fie

ca populația combinată A și B. Alegerea populației standard nu contează. Expresia

„populația standard” în acest context se referă la o populație cu o anumită distribuție de vârstă (dacă ne ajustăm pentru vârstă) sau distribuție pe sexe (dacă ne ajustăm pentru sex). Ratele specifice vârstei (sau specifice sexului, dacă acesta este ceea ce este ajustat) pentru ambele grupe A și B sunt aplicate distribuției de vârstă a populației standard pentru a compara A și B **ca și cum** ar avea aceeași distribuție de vârstă. .

Rețineți că dacă utilizați două populații standard diferite, obțineți rate diferite ajustate în funcție de vârstă, dar cifrele relative sunt aceleași, adică ratele ajustate în funcție de vârstă pentru A sunt mai mici decât pentru B. Aceasta înseamnă că ratele specifice vârstei pentru A sunt mai mici. decât pentru B, dar deoarece ratele brute sunt aceleași, trebuie să însemne că populația A este mai în vârstă. Deoarece știm că ratele specifice vârstei pentru persoanele în vârstă sunt mai mari decât pentru persoanele mai tinere, populația A trebuie să fi fost ponderată de o proporție mai mare de persoane în vârstă (care au contribuit la mai multe decese) pentru a avea ca rezultat aceeași rată brută ca B, dar în o rată mai mică ajustată în funcție de vârstă.

Există excepții de la concluzia de mai sus atunci când luăm în considerare grupurile în care mortalitatea infantilă este foarte mare. În acest caz, s-ar putea ca tinerii să aibă rate de deces foarte mari, chiar mai mari decât bătrânii. În societățile industrializate, totuși, ratele de deces specifice vârstei pentru bătrâni sunt mai mari decât pentru tineri.

POPULAȚIA STANDARD I: Exemplu (mai mulți bătrâni decât tineri)

Vârstă	# de oameni	Aplicați ratele de deces specifice vârstei pentru populația A la populația standard	# de decese așteptate în A dacă ar fi aceeași compoziție de vârstă ca populația standard	Aplicați ratele de deces specifice vârstei pentru populația B la populația standard	# de decese așteptate în B dacă ar fi aceeași compoziție de vârstă ca populația standard
Tineri	300 ×	.004 =	1.2	.005	1.5
Vechi	700 ×	.016 =	11.2	.020	14.0
Total	1.000		12.4		15.5
Tarife ajustate în funcție de vârstă pentru: A = 12/1.000 B = 15/1.000					

POPULAȚIA STANDARD II: Exemplu (mai mulți tineri decât bătrâni)

Vârstă	# de oameni	Aplicați ratele de deces specifice vârstei pentru populația A la populația standard	# de decese așteptate în A dacă ar fi aceeași compoziție de vârstă ca populația standard	Aplicați ratele de deces specifice vârstei pentru populația B la populația standard	# de decese așteptate în B dacă ar fi aceeași compoziție de vârstă ca populația standard
Tânăr	1.167 ×	.004 =	4.67	.005	5.84
Vechi	833 ×	.016 =	13.33	.020	16.66
Total 2.000			18		22,50

Tarife ajustate în funcție de vârstă pentru:	Tarife ajustate în funcție de vârstă pentru:
$\frac{18}{\text{A } 2.000}$	$\frac{22 \,.\, 5}{\text{B } 2.000}$
$= \frac{9}{1.000}$	$= \frac{11 \,.\, 25}{1.000}$

Appendix 6
DETERMINAREA ADECVĂRII SCORURILOR DE MODIFICARE

(1) Pentru a determina dacă scorurile de modificare sunt adecvate:

Luați în considerare un grup de 16 pacienți care au următoarele scoruri pe o scală care evaluează simptomele depresive; un retest este dat la scurt timp după aceea pentru a determina variabilitatea scorurilor în cadrul pacienților (Tabelul A.4):

Tabelul Testați și scoruri

# pacient	Primul Test Scala Scor	Retestați Scor
1	12	13
2	16	15
3	22	21
4	24	23
5	30	29
6	18	19
7	16	15
8	12	12
9	14	15
10	18	18
11	24	24
12	30	29
13	18	19
14	16	15
15	14	15
16	10	11

Medie	18.38	18.31

O analiză a varianței indică următoarele (Tabelul A.5):

Tabelul A.5 Analiza varianței

Sursa de variație	SS	df	DOMNIȘO	F	*P* -
Pacienții	1014,7188	15	67,6479	156,8647	.0000
Testare retest	.0313	1	.0313	.0725	.7915
Eroare	6,4688	15	.4313		
Total	1021.2188	31			
σ între pacienți	(67,65 – .43)/2 = 33,61				
σ între pacienți + e	; rror = 33,61/(33,61 + .43) = .987				

Acesta este mai mare de 0,5, astfel încât utilizarea scorurilor de modificare este adecvată . [Notă: σ^2, sau varianța, este MS (pătrat mediu) din analiza varianței.]

În continuare, pacienții sunt împărțiți în două grupe; unui grup i se administrează o intervenție alimentară care durează 10 săptămâni, în timp ce celălalt grup servește ca grup de control. Scala este administrată din nou după 10 săptămâni la ambele grupuri, cu următoarele rezultate (Tabelul A.6):

(2) Pentru a calcula **coeficientul de sensibilitate** la schimbare, faceți o analiză a varianței cu măsuri repetate asupra scorurilor din **grupul de tratament** ; pentru a obține varianța erorii, calculați varianța scorurilor de modificare (Tabelul A.7).

Coeficient de sensibilitate = varianța scorurilor de modificare în grupul de tratament/(varianța scorurilor de modificare + varianța de eroare) =

$$\frac{44 : 57}{(44{:}57 + 22{:}57)} = 66$$

(44,57 este obținut din ultima coloană a tabelului A.6).

(3) *Mărimea efectului* = media scorurilor de modificare /sd a scorurilor pretest: în

grupul de tratament = —11/6,32 = —1,74 (a existat o *scădere a scorului simptomelor depresive* de 1,74 unități de abatere standard pretest).

Tabelul A.6 Schimbarea scorurilor între pre și post-test

Grupul de control				Grup de tratament			
# pacient	Pretest	Post test	Schimbați scorul	# pacient	Pretest	Post test	Schimbați scorul
1	12	13	1	9	14	7	–7
2	16	15	–1	10	18	10	–8
3	22	20	–2	11	24	7	–17
4	24	18	–6	12	30	5	–25
5	30	25	–5	13	18	10	–8
6	18	16	–2	14	16	8	–8
7	16	12	–4	15	14	4	–10
8	12	10	–2	16	10	5	–5
Medie	18.75	16.13	–2,63		18	7	–11
Varianta	38,79	23.27	5.13		40.00	5.14	44,57
sd	6.23	4,82	2.26		6.32	2.27	6,68

Tabelul A.7 Analiza varianței scorurilor modificării

Sursa de variație	SS	df	DOMNIȘOARĂ	F	Valoare
Între test/retest	484	1	484,0000	21,4430	.0004
In cadrul pacientilor	316	14	22,5714		
Total	800	15			

(4) **Măsura de răspuns a lui Guyatt** [30] este (scorurile medii ale modificării în grupul de tratament)/(sd a scorurilor modificării la subiecții stabili). Presupunem aici că grupul de control este grupul de subiecți stabili, deși în general „subiecți stabili” se referă la subiecții stabili în raport cu un

anumit criteriu extern.

$$G \quad \begin{matrix} -11 \\ \\ 2{:}26 \end{matrix} \quad -4 : 86$$

(5) **Comparație cu un grup de control:** Mărimea efectului pentru grupul de tratament este de —1,74, deci depășește în mod clar modificarea grupului de control , care este —2,63/6,23 = —.42. Dacă calculăm raporturile dintre tratament și grupul de control pentru indicii de receptivitate de mai sus, vom constata în acest exemplu că aceștia sunt foarte similari.

Pentru mărimea efectului, raportul este -1,74/-.42 = 4,14. Pentru statistica lui Guyatt este -4,86/-1,16 = 4,19. (—1,16 a fost obținut prin modificarea medie în grupul de control împărțit la abaterea standard a scorurilor de schimbare în grupul de control, adică = —2,63/2,26.)

Pentru coeficientul de sensibilitate, acesta este .66/.14 = 4.71. (Valoarea 0,14 a fost obținută făcând o analiză a scorurilor din grupul de control, care nu este prezentată aici, așa că luați-l pe credință sau calculați-l ca o verificare a acurateței acesteia.)

REFERINȚE

Capitolul 1

1. Popper KR: Logica descoperirii științifice. New York: Harper and Row, 1959.
2. Weed DL: Despre logica inferenței cauzale. Jurnalul American de Epidemiologie, 123(6):965–979, 1985.
3. Goodman SN, Royall R: Dovezi și cercetare științifică. AJPH, 78(12):1568–1574, 1988.
4. Susser M: Reguli de inferență în epidemiologie. În: Regulatory Tox icology and Pharmacology, Capitolul 6, pp. 116–128. New York: Academic Press, 1986.
5. Brown HI: Percepție, teorie și angajament: noua filosofie a științei. Chicago: Precedent, 1977.
6. Hill AB: și Hill ID, Bradford Hill 's Principles of Medical Statistics , Hodder Education Publishers, 1991
7. Feinstein AR, Principiile statisticii medicale, New York; Chapman și Hall/CRC, 2001

Capitolul 3

8. Drapkin A, Merskey C. Terapia anticoagulantă după infarctul miocar acut. Relația dintre beneficiul terapeutic și vârsta pacientului , sexul și severitatea infarctului. **Jurnalul Asociaţiei Medicale Americane** . 1972;222:541–548
9. Dezvoltarea intelectuală a copiilor. Departamentul de Sănătate, Educație și Bunăstare al SUA, Serviciul de Sănătate Publică. HSMA, Seria de statistici vitale și de sănătate 11, 1971
10. Davis BR, Blaufox MD, Hawkins CM, Langford HG, Oberman A, Swencionis C, Wassertheil-Smoller S, Wylie-Rosett J, Zimbaldi N. Trial of antihipertensive interventions and management. Proiecta,

    metode și rezultatele de referință selectate. **Studii clinice controlate** . 1989;10:11–30
11. Oberman A, Wassertheil-Smoller S, Langford HG, Blaufox MD, Davis BR,

Blaszkowski T, Zimbaldi N, Hawkins CM. Tratamentul farmacologic și nutrițional al hipertensiunii ușoare: modificări ale stării de risc cardiovascular. **Analele medicinei interne** . 1990;112:89–95

12. Rothman KJ. Nu sunt necesare ajustări pentru comparații multiple. **Epidemiologie** . 1990;1:43–46
13. Scarr-Salapatek S. Rasa, clasa socială și iq. **Știința** . 1971;174:1285–1295
14. Sokal RR RJ. **Biometrie** . Ediția a 4-a, San Francisco: WH Freeman; 2011
15. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Analiza statistică a datelor corelate folosind ecuații de estimare generalizate: O orientare. **Jurnal american de epidemiologie** . 1971.

Capitolul 4

16. Grupul de Cooperare pentru Programul de Detectare și Urmărire a Hipertensiunii, Studii ale tensiunii arteriale în 14 comunități. Un ecran în două etape pentru hipertensiune arterială. **Jurnalul Asociației Medicale - Americane** . 1977;237:2385–2391
17. Sorlie PD, Allison MA, Aviles-Santa ML, Cai J, Daviglus ML, Howard AG, Kaplan R, Lavange LM, Raij L, Schneiderman N, Wassertheil-Smoller S, Talavera GA. Prevalența hipertensiunii, conștientizarea, tratamentul și controlul în studiul/studiul asupra sănătății comunității hispanice asupra latinilor. **Jurnalul american de hipertensiune arterială** . 2014;27:793–800
18. Inter-Society Commission for Heart Disease Resources, Raport al comisiei inter-societate pentru resursele bolilor de inimă. Prevenirea primară a bolilor aterosclerotice. **Circulația** . 1972; XLII:1–44
19. Kannel WB, McGee D, Gordon T. Un profil general de risc cardiovascular: Studiul Framingham. **Jurnalul american de cardiologie** . 1976;38:46–51
20. Aceste date provin din Proiectul Național de Pooling. În scopul acestui exemplu, tensiunea arterială ridicată este definită ca tensiune arterială diastolică > = 105 mmHg. Boala în cauză este un „Eveniment coronarian” și perioada de timp este de 10 ani . Rețineți că hipertensiunea arterială este definită în prezent ca tensiune arterială sistolică > =140 mmHg și/sau tensiune arterială diastolică > =90 mmHg
21. Lilienfeld AM, Lilienfeld D,. Foundations of Epidemiology, New York,

Chapel Hill, NC, 1980

22. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J. Riscurile și beneficiile estrogenului plus progestativ la femeile sănătoase în postmenopauză: rezultate principale din studiul controlat randomizat al inițiativei pentru sănătatea femeilor . **JAMA: jurnalul Asociației Medicale Americane** . 2002;288:321–333.
23. Exemplul numeric este oferit de Dr. Martin Lesser, Centrul Medical al Universității Cornell, New York
24. D ' Agostino RB, Jr. Metode de scor de propensie pentru reducerea părtinirii în comparația unui tratament cu un grup de control nerandomizat. **Statistica în medicină** . 1998;17:2265–2281
25. Smoller JW, Allison M, Cochrane BB, Curb JD, Perlis RH, Robinson JG, Rosal MC, Wenger NK, Wassertheil-Smoller S. Utilizarea antidepresivelor și riscul de morbiditate și mortalitate cardiovasculară incidentă în rândul femeilor în postmenopauză în studiul inițiativă pentru sănătatea femeilor . **Arhivele medicinei interne** . 2009;169:2128–2139
26. Greenland S. Modelarea și selecția variabilelor în analiza epidemiologică. **Jurnalul American de Sănătate Publică.** 1989;79:340–349
27. Wassertheil-Smoller S, Fann C, Allman RM, Black HR, Camel GH, Davis B, Masaki K, Pressel S, Prineas RJ, Stamler J, Vogt TM. Relația dintre masa corporală scăzută și deces și accident vascular cerebral în programul hipertensiunii sistolice la vârstnici. Grupul de cercetare cooperativă ovină. **Arhivele medicinei interne** . 2000;160:494–500

Capitolul 6

28. Grupul de cercetare al proiectului de medicamente coronariene: Influența aderenței la tratament și răspunsul colesterolului asupra mortalității în proiectul de medicamente coronariene. **New England Journal of Medicine,** 303:1038–1041, 1980.
29. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Proiectarea și analiza studiilor clinice randomizate care necesită observarea prelungită a fiecărui pacient. I. Introducere și Proiectare. **British Journal of Cancer,** 34:585–612, 1976.

Capitolul 7

30. Guyatt G, Walter S, Norman G: Măsurarea schimbării în timp: evaluarea utilităţii instrumentelor de evaluare. J Chronic Dis 1987;40:171–8.

Capitolul 8

31. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică. I. Concepte de bază în genetica umană și tehnologia de laborator. Analele epidemiologiei. 1999;9:1–16
32. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică ii. Probleme în proiectarea studiului și cartografierea genelor. Analele epidemiologiei. 1999;9:75–90
33. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică iii. Bioinformatica si metode genetice statistice. Analele epidemiologiei. 1999;9:207–224
34. Duncan LE, Pollastri AR, Smoller JW. Luați în considerare decalajul: de ce mulți geneticieni și oameni de știință în psihologie au opinii discrepante despre cercetarea interacțiunii gene-mediu (gxe). Psihologul american . 2014;69:249–268
35. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts

de gemeni din Suedia, Danemarca și Finlanda. Jurnalul de medicină din New England. 2000;343:78–85
36. Hinrichs AL, Larkin EK, Suarez BK. Stratificarea populației și modele de dezechilibru de legături. Epidemiologie genetică. 2009;33 Suppl 1:S88-92
37. Corvin A, Craddock N, Sullivan PF. Studii de asociere la nivel de genom : Un primer. Medicina psihologica. 2010;40:1063–1077
38. Chen JJ, Roberson PK, Schell MJ. Rata de descoperire falsă: un concept cheie în studiile genetice la scară largă. Controlul cancerului: jurnalul Moffitt Cancer Center. 2010;17:58–62

Sham PC și Purcell SM. Puterea statistică și testarea semnificației în studii genetice la scară largă. Nat Rev Genet. 2014;15:335–46.

Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S și Sunyaev S. Studii de secvențiere în genetica umană: proiectare și interpretare. Nat Rev Genet. 2013;14:460–70.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ și Sham PC. PLINK: un set de instrumente pentru asocierea întregului genom și analize de legături bazate pe populație . Am J Hum Genet. 2007;81:559–75.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP și Hirschhorn JN. Studii de asociere la nivelul genomului pentru trăsături complexe: consens, incertitudine și provocări. Nat Rev Genet. 2008;9:356–69.

Price AL, Zaitlen NA, Reich D și Patterson N. Noi abordări ale stratificării populației în studiile de asociere la nivel de genom. Nat Rev Genet. 2010;11:459–63.

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR și Lander ES. Căutarea eredității lipsă: proiectarea unor studii de asociere cu variante rare. Proc Natl Acad Sci US A. 2014;111:E455-64.

Capitolul 9

39. Kannel WB, McGee D, Gordon T. Un profil general de risc cardiovascular: Studiul Framingham. **Jurnalul american de cardiologie** . 1976;38:46–51

40. D ' Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Profil de risc de accident vascular cerebral: ajustare pentru medicamente antihipertensive. Studiul Framing ham. **Accident vascular cerebral; un jurnal de circulatie cerebrala** . 1994;25:40–43

41. Anderson KM, Odell PM, Wilson PW, Kannel WB. Profiluri de risc pentru bolile cardiovasculare. **Jurnalul american al inimii** . 1991;121:293–298

42. Wassertheil-Smoller S, McGinn A, Allison M, Ca T, Curb D, Eaton C, Hendrix S, Kaplan R, Ko M, Martin LW, Xue X. Îmbunătățirea predicției riscului de accident vascular cerebral: Rolul proteinei c-reactive și al lipoproteinelor- fosfolipaza a(2) asociată în inițiativa pentru sănătatea femeilor . **Jurnalul Internațional de AVC: jurnalul oficial al Societății**

Internaționale de AVC . 2012

43. Pencina MJ, D ' Agostino RB, Vasan RS. Metode statistice pentru evaluarea utilității suplimentare a noilor biomarkeri. **Chimie clinica si medicina de laborator : CCLM / FESCC** . 2010;48:1703–1711
44. Pencina MJ, D ' Agostino RB, Sr., Steyerberg EW. Extensii ale calculelor de îmbunătățire a reclasificării nete pentru a măsura utilitatea noilor biomarkeri. **Statistica în medicină** . 2011;30:11–21
45. Pencina MJ, D ' Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpretarea valorii incrementale a markerilor adăugate la modelele de predicție a riscului. **Jurnal american de epidemiologie** . 2012;176:473–481

LECTURI SUGERATE

Anderson S, Auquier A, Hauck WW, Oakes D, Vandaele W, Weisberg HI: Statistical Methods for Comparative Studies. New York: John Wiley and Sons, 1980.[4]

Cohen J: Analiza puterii statistice pentru științe comportamentale, ed. a 2-a. Hillsdale, NJ: Lawrence Erlbaum, 1988.[5]

Cohen J (ed.), Cohen P, West SG, Aiken LS: Applied Multiple Regression: Corelation Analysis for the Behavioral Science, ed. a 3-a, Lawrence Erlbaum, 2002[6]

Elwood JM: Relații cauzale în medicină. Un sistem practic de evaluare critică. New York: Oxford University Press, 1988.[7]

Fleiss JL, Levin B, Paik MC: Statistical Methods for Rates and Proportions , ed. a 3-a. New York: John Wiley and Sons, 2003[8]

Friedman, LM, Furberg CD, DeMets DL, Fundamentals of Clinical Trials, ediția a 4-a, Springer, 2010[9]

Haynes BR, Sackett DL, Gordon H. Guyatt GH, Tugwell P Clinical Epidemiology: How to Do Clinical Practice Research, ediția a 3-a, LWW, 2005

Hill, B: Metode statistice în medicina clinică și preventivă. Edinburgh și Londra: ES Livingstone, 1962.[10]

Hosmer DW, Lemeshow S, Sturdivant RX: Regresia logistică aplicată. Ediția a

[4] Nivel intermediar. Necesită un curs anterior de statistică. Excelent pentru cercetătorii aplicați.

[5] Un clasic și o necesitate. Tot ce ați dorit vreodată să știți despre calculele dimensiunii eșantionului, explicat clar și cuprinzător. O carte de referință și sursă. Probabil trebuie interpretat de un statistician.

[6] Expunere excelentă, amănunțită și clară a subiectelor foarte complexe. E destul de avansat.

[7] Carte excelentă, cu explicații clare despre modelele de studiu, conceptele epidemiologice și metodele statistice relevante.

[8] Un excelent text clasic de statistică de nivel doi, care se referă la analiza datelor calitative sau categoriale.

[9] O carte excelentă a 3 experți renumiți în domeniu.

[10] Text clasic pe subiect.

Hosmer DW, Lemeshow S: Applied Logistic Regression. New York: John Wiley and Sons, 2013.[11]

Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH: Biostatistics in Clinical Medicine, ediția a 3-a,. New York: Macmillan, 1993.[12]

Kelsey JL, Whittmore AS, Thompson WD, Evans AS, Thompson: Methods in Observational Epidemiology. New York: Oxford University Press, 1996.[13]

Kleinbaum DG, Kupper LL Inizam A, Muller KE,: Analiza de regresie aplicată și alte metode multivariabile. Ediția a patra, Thomson Books/Cole 2007.[14]

Meinert CL Clinical Trials: Design, Conduct and Analysis (Monografii în Epidemiologie și Biostatistică) în colaborare cu Tonascia S. Ediția a doua, New York, Oxford: Oxford University Press, 2012.

Rothman KJ, Lash TL și Greenland S: Epidemiologie modernă, Ed. a 3-a. Philadelphia: Lippincott Williams & Wilkins, 2012.[15]

Siegel S și Castellan Jr. NJ: Statistici nonparametrice pentru științe comportamentale. Ediția a II-a. New York, Toronto, Londra: McGraw-Hill, Inc., 1988.[16]

Sokal RR, Rohlf JF: Biometrie, ed. a 4-a. San Francisco: WH Freeman, 2011.[17]

Streiner L, Norman GR: Scale de măsurare a sănătății. Un ghid practic pentru Dezvoltarea și utilizarea lor, ediția a IV-a. New York: Oxford University Press, 2008.[18]

Szklo M, Nieto FJ: Epidemiologie. Dincolo de elementele de bază. ediția a 3-a,

[11] Carte avansată despre construirea de modele în regresia logistică, necesită cunoștințe statistice.

[12] Adresat mai ales medicilor. Nivel intermediar. Foarte bune explicații ale biostatisticii. Multe exemple din literatura de specialitate.

[13] Explicații extrem de utile ale problemelor implicate în studiile caz-control, retrospective și prospective. O bună discuție despre potrivire, stratificare și probleme de proiectare.

[14] Acest text, după cum indică titlul, tratează regresia multiplă și subiecte conexe.

[15] Un text avansat care acoperă atât problemele de design, cât și cele statistice. Accentul este un studiu epidemiologic observațional și este îndreptat mai mult către cercetător decât către clinician.

[16] O carte „cum se face”; o referință excelentă; conturează fiecare procedură pas cu pas, un clasic.

[17] Lucrări de referință cuprinzătoare.

[18] Carte excelentă și de înțeles despre construcția scalei și schimbarea scorurilor.

Jones & Bartlett Learning; 2012.[19]

[19] Undeva între textul de nivel întâi și al doilea, care se distinge prin stilul său clar și foarte lizibil, și multe exemple excelente.

INDEX

Risc absolut, 98, 103, 126, 127, 213
Precizia testului, 134
Legea aditivă a probabilității, 45
Efecte de interacțiune a modelului aditiv, 128–129
Valoarea aditivă a unui biomarker, 206–213
Rate ajustate în funcție de vârstă, 86–88, 234, 235
Ajustare de vârstă, 87, 88, 98, 233–235
Rata anuală de mortalitate specifică vârstei, 86
Alele, 174, 176, 178–192, 195, 196–199, 202
Abordarea partajării alelelor, 180
Eroare alfa, 15
Ipoteza alternativă, 11, 14, 57, 58, 60, 61, 147, 148, 180
Analiza variației (ANOVA), 63–66, 68, 70, 71, 81, 167, 238
  cu doi factori (vezi ANOVA cu doi factori)
ANOVA. Vezi Analiza variației
Afirmații, 4, 5, 13
Asociația, 8, 9, 71, 72, 74, 99, 108, 119, 123, 179, 181–183, 185, 186–190, 195–198, 201, 202
Asociere și cauzalitate, 71–73
Studii de asociere, 174, 176–179, 180–183, 188–190, 195
Risc atribuibil (AR), 103–104, 128
Politicile de autor, 223
Perechi de baze, 172, 173, 178
Probabilitate bayesiană, 21–22
Curbă în formă de clopot, 37, 71
Beneficență, 219
Eroare beta, 16
Varianta între grupuri, 64
Procedura Bonferroni, 66–67, 71, 161

Gene candidate, 178, 182–183, 189, 190
Studiu caz-control, 8, 9, 92, 98–101, 108, 109, 111, 181, 185, 192, 197, 206–208, 216
  calculul riscului relativ din, 99–103
Variabile categoriale, 10, 57, 175
NRI fără categorii, 213–214, 216
Căi cauzale, 75–76
Rata anuală de mortalitate specifică cauzei, 85
Diviziunea celulară, 187
Centimorgan (cM), 177, 178
Tendința centrală, 34, 35
Schimbați scorurile, 163, 167, 168, 237–240
  determinarea caracterului adecvat al, 237–240
Chi-pătrate, 19, 27–33, 56, 57, 81, 126, 180, 185, 186, 192–195, 225, 229, 230
  valori critice de, 30, 33, 225
Testul chi-pătrat, 19, 27, 28, 31, 33, 56, 57, 81, 192
Localizare cromozomală, 177
Semnificație clinică, 155
Studiile clinice ca „standard de aur”, 123, 148–149, 222
  randomizat, 122, 123, 143–145, 148
  dimensiunea de, 122, 149
Coeficient de sensibilitate, 167, 238, 240

Cohorta, 9, 90, 93–97, 122, 197
Studiu de cohortă, 9
Comparații între două grupuri, 54
Ratele de concordanță, 175
Probabilitate condiționată, 20–21, 24, 25 Intervale de încredere
în jurul diferenței dintre două medii, 53–54
pentru proporții, 52–53
Limite de încredere, 50–53, 77, 102, 103, 116 privind raportul de cote, 102
Confuzie prin indicație, 117–118, 120
Variabile de confuzie, 107–108, 114
Conjecturi și respingeri, metoda, 5
Consimțământ, informat, 219–221
Tabelul de urgență, 27, 29
Variabile continue, 10, 27, 34, 52, 57, 63, 79, 116, 175
Grupul de control, 27–29, 53, 55, 56, 99, 109, 115, 122, 143–147, 150, 152, 156–159, 163, 168, 183, 227, 238–240
Calcularea coeficientului de corelație, 231–232
legătura dintre regresia liniară și, 79
Modelul de pericole proporționale Cox, 114–116, 123, 126, 129, 130, 146, 206, 216
Valoare critică
de chi-pătrat, 30, 33, 225
din Z, 56, 225
lui Cronbach , 165
Studiu transversal, 8, 90 studiu longitudinal versus , 93–96
Rata brută anuală a mortalității, 85 C-statistică, 120, 215, 216
Punct de limită, 45, 140–142, 150, 155

Consiliul de monitorizare a datelor și siguranței (DSMB), 149, 224
Rata mortalității, 11, 85–87, 130, 233–235
Inferență deductivă, 1, 4–6
Grade de libertate (df), 50, 52, 54, 64, 65, 67, 126, 225, 229, 230
Acid dezoxiribonucleic (ADN), 172, 177, 181, 189, 190, 196, 200, 201
Variabile dependente, 76, 79, 92, 100, 107, 111, 114, 115, 119, 121, 125, 129, 170
Epidemiologie descriptivă, 83 df. Vezi Grade de libertate (df) Diferența dintre două înseamnă intervale de încredere în jur, 53–54
calculul dimensiunii eșantionului pentru testare, 155–159, 161
eroare standard de, 42–43, 46, 49, 58, 59, 215
testul t pentru, 57–58
Diferența dintre două proporții, calculul mărimii eșantionului pentru, 159–160
Diferențele, distribuția, 42
Variabile discrete, 10, 27, 34
Distribuția diferențelor, 42
ADN. Vezi Acid dezoxiribonucleic (ADN)
Studiu dublu-orb, 144, 152

Dimensiunea efectului, 155–159, 168, 170, 176, 196, 197, 202, 238, 240
Epidemiologie
definit, 83
descriptiv, 83, 84
impactul genetic (vezi Epidemiologia genetică) asupra

vieților umane, 241–242 molecular, 172
utilizări ale, 83–84
Echilibrul, 222
Eroare pătrat mediu, 66, 70
Erori, tipuri de, 13–15, 81, 157
Etică, cercetare, statistică și, 217–224
Etnie, 119, 185
Exoni, 202
Frecvența așteptată, 18, 28, 29, 31
Studii experimentale, 8, 9, 83, 92

Notație factorială, 227
Rata fals-negative, 134, 135
Fals-negative, 140, 141, 196–197
Rata fals-pozitive, 133–135, 139
Fals pozitive, 140, 141, 142, 183, 186, 190, 195–198
Falsificare, 5
Testul exact al lui Fisher, 31 , 227–228
Raportul F, 64–66
Frecvență, așteptată, 18, 28, 29, 31
Genele, 173–180, 182–183, 189, 190, 202–204
Arhitectura genetică, 176
Epidemiologie genetică, prezentare generală a, 173–174
Analiza legăturii genetice, 177
principii genetice,
Genomul, 172, 173, 177, 178, 185, 189, 190, 195, 197, 200–203
Secvențierea genomului, 200–201
Studii de asociere la nivelul genomului (GWAS), 174, 183, 189–202
Controlul genomic, 185
Genomica, 172, 204
Măsura de răspuns a lui Guyatt , 239–240

HapMap, 189
Hardy–Weinberg, 190–192
Raport de risc, 99, 115, 116, 119, 121–124, 130, 153
Indicele de ereditabilitate, 175, 176
Eterogenitate, 186, 188
Proiectul genomului uman, 201, 203
Vieți umane, impactul epidemiologiei asupra, 241–242
Subiecți de cercetare umană, protecția, 218–220
Metoda ipotetico-deductivă, 2

Rata de incidență, 89, 91, 97, 100, 126
Variabile independente, 68, 69, 76, 79, 80, 92, 100, 107, 109, 110, 115, 118, 119, 170
Inferență inductivă, 2–4, 6–8
Inferență
deductiv, 1, 4–6
inductiv, 2–4, 6–8
Consimțământ informat, 219–221
Discriminare integrată Îmbunătățire (IDI), 214–215
Indicele de discriminare integrat, 214–215
Integritate, cercetare, 222–223
Analiza intenției de a trata, 152–153
Interacțiune(e), 68–70, 125–129, 182
aditiv, 125–129
între două variabile, 69
multiplicative, 125–129

Probabilitate comună, 19, 28
în formă de J, 129–132
Justiție, 219, 220

Curbele de supraviețuire Kaplan-Meier, 112
Kappa, 33–34, 167
Testul Kruskal–Wallis, 71

Lambda (λ), 185
Potrivirea celor mai mici pătrate, 77
Metode tabelului de viață, 112–114
Raportul de probabilitate (LR)
a unui test negativ, 135
a unui test pozitiv, 134
Regresia liniară
legătura dintre și coeficientul de corelare, 79
multiplu, 79–80
Legătura
dezechilibru, 186, 189, 190, 195, 198
statistică, 179–180 studii, 174, 178, 182
Scorul LOD, 179–180
Regresie logistică, multiplă, 98, 109–112, 115, 119, 182
Studiu longitudinal, 9, 80, 149 studiu transversal versus , 93–96
LR. Consultați Raportul de probabilitate

Complot Manhattan, 197–198
Testul t perechii potrivite, 60–61
Potrivire, 108–109, 118, 121, 201
Testul McNemar, 31–33
Mijloace)
diferenta intre doua (vezi Diferența dintre două mijloace)
regresia spre, 149–151
eroare standard de, 39–42, 46
Pătrat mediu, 65, 66, 70, 238
Median, 34, 186, 194
Meioza, 177, 179
Mod, 34, 180, 196, 197
Epidemiologie moleculară, 172
Morbiditate, 88, 90, 97, 104
Rata mortalității, 19, 27, 84–87, 90, 152
Regresie liniară multiplă, 79–80
Regresie logistică multiplă, 98, 109–112, 115, 119, 182
Legea multiplicativă a probabilității, 21, 24
Interacțiunea modelului multiplicativ efecte, 125–129
Modele multivariate, selectând variabile pentru, 123–125
Mutații, 181, 201, 202
Evenimente care se exclud reciproc, 18–20

Nadir al unui pătratic relație, 131
Test negativ
raportul de probabilitate a, 135
valoarea predictivă a a, 134, 136
Îmbunătățirea netă a reclasificării, 212–214
Relații neliniare, 129–132
metode de legare neparametrică, 180
Test neparametric, 71, 229–230
Indian nerezident (NRI), 212–214, 216
Distribuție normală, 38, 40, 41, 71
standardizat, 44–47
Ipoteza nulă, 6, 10–15, 27, 30, 32, 33, 55–58, 60–64, 66, 117, 126, 128, 147, 148, 154, 155, 169, 180, 186, 193, 215, 225, 228, 230
testare, 8, 10, 11, 14

Studii observaționale, 3, 8, 92, 117, 118, 122, 148, 149, 208, 219, 222
Observații, 1–8, 12, 15, 37, 80, 83, 89–91, 115, 228

Raportul de cote (OR), 23, 99–103, 105,
110, 111, 123, 182, 192, 196, 197, 199
limitele de încredere asupra, 102
Test cu o singură coadă, 58, 148, 157, 225
test cu două cozi versus , 147–148, 157
Potrivire unu-la-unu, 108
SAU. Vedeți raportul de cote (OR)
Intervale de confidențe suprapuse, 117

PAR. A se vedea Riscul atribuibil populației (PAR)
Analiza legăturii parametrice, 180
Percepția pacientului , 163, 170
Pătrunderea, 180
Persoană-ani de observație, 90–91
Fenotip, 175, 176, 181, 182, 188, 189, 196, 198
Placebo, 3, 62, 65, 67, 68, 70, 79, 104, 115, 143–145, 147, 149, 150, 152, 153, 168, 169, 217, 218, 221, 222
Prevalența punctuală, 88
Scorul Polygene, 198–200
Polimorfisme, 172, 181, 182, 189
Estimare combinată, 43, 56, 59, 65, 66, 70
Populația
stratificare, 183–186, 194, 195
valori, 41, 48
Riscul atribuibil populației (PAR), 103, 104
Test pozitiv, 133–137, 139
raportul de probabilitate a, 134
valoarea predictivă a lui a, 134–137
Probabilitatea de îmbolnăvire posttest, 135
Putere, statistică, 156
Valoare predictivă
a unui test negativ, 134, 136
a unui test pozitiv, 134–137
Probabilitatea de îmbolnăvire pretest, 135
Rata de prevalență, 22, 24, 88–90, 103, 104, 135–139, 184, 185, 196, 197, 207
Model probabilistic, 6
Probabilitate(e)
Bayesian, 21–22
combinând, 18–20, 61, 63
condiționat, 20–21, 24, 25
de boală, pretest și posttest, 135
comun, 19, 28
Înclinația
analiză, 118–123
potrivire, 118, 121
Proporții
intervale de încredere pentru, 52–53
diferența dintre două, calculul mărimii eșantionului pentru, 159–160
erori standard ale, 74
Testul Z pentru compararea a două, 55–56 Studii prospective, 9, 92, 93, 96–98, 100, 206, 208
calculul riscului relativ din, 98–99
Proteomică, 172
Psihometrie, 166, 169
p , 66, 67, 124, 130, 154, 155, 192–194, 196, 197, 199, 213, 239

Quadratic, 130, 131
Cuantile după diagrame cuantile (parcele QQ), 192–195

Calitatea vieții
nevoia de standarde de cercetare în, 170
capcane în evaluare, construcția la scară 169–170 pentru evaluare, 164

Randomizare, scopuri de, 145–146
Atribuire aleatorie, performanță, 146–147
Studii clinice randomizate, 123, 143–145
Interval de date, 37
Variante rare, 200–202
Rata, 2, 3, 11, 19, 27, 77, 84–91, 96, 97, 100, 104, 105, 112, 114, 115, 126, 130, 133–135, 139, 139, 15–12 159, 172, 175, 196, 197, 228, 233–235
Curba caracteristicii de funcționare a receptorului (ROC), 142, 215
Reclasificare, 212–214 Recombinare, 177, 179, 180 Referințe, 6, 74, 88, 125, 126, 174, 180, 190, 193, 194, 197
Regresia
coeficienți, 79 calcul, 231–232
liniile, 76, 78, 79
spre medie, 149–151
Risc relativ (RR)
calculul, din studii prospective, 98–99
estimare a, din studii caz-control, 99–103
măsuri de, 96–98
Fiabilitate, 8, 164–166 Repetabilitate, 164
Etica cercetării, statistică și, 217–224
Integritatea cercetării, 222–223
Cercetare, necesitatea standardelor de evaluare a calității vieții, 237, 169, 170
Subiecte de cercetare, uman, protecția, 218–220
Respectul pentru persoane, 219
Prejudecăți de răspuns, 105–107
Reactivitate, 164, 167–168, 240 Studii retrospective, 8, 9, 100
Risc
factori, 89, 97, 99, 101, 103–105, 120, 122–124, 145, 171, 172, 176, 199, 200, 205, 210, 242
predicție, 205–216
RR. Vezi Risc relativ

Calculul dimensiunii eșantionului
pentru diferenta intre doua proporții, 159–160
pentru a testa diferența dintre două mijloace, 160–161
Valori eșantion, 48
Metoda științifică, 1–16, 92
Raționamentul științific, logica, 1–6
Screening, 88, 90, 133–142, 150
Selectarea variabilelor pentru multivariate
modele, 123–125
Auto-raport, 163
Sensibilitate, 133–142, 167, 215, 238, 240
Secvențiere, 200–202
Nivel de semnificație, 13, 15, 62–64, 66, 124, 156, 157, 159–161, 195, 197, 225, 228
Studiu single-orb, 144
Polimorfisme cu un singur nucleotide (SNP), 173, 181, 183–186, 189–200, 202–204
Formula Spearman-Brown, 165
Specificitatea, 133–142

Abaterea standard
diferența dintre standard eroare și, 39–42
sensul lui, 37–38
Eroare standard
diferența dintre standard abatere și, 39–42
de diferenta intre doi înseamnă, 42–43, 58
de medie, 40, 42, 46
de proporție, 52, 53
Rata standardizată de mortalitate (SMR), 90
Distribuție normală standardizată, 44–47
Scor standard, 44
Metode statistice, 7, 27, 31, 146, 172
Puterea statistică, 156
Semnificație statistică, 74, 154, 155, 196
Statistici, 6–8, 18, 27–81, 90, 171, 180, 185–187, 192–195, 217–224
etica cercetării și, 217–224
Berze și bebeluși, 74
Studii
proiectarea, 8–9, 13, 155
tipuri de, 8, 92–93
Lecturi sugerate, 61, 108, 125, 146, 225, 230
Analiza supraviețuirii, 112–114
Curbele de supraviețuire, 112, 114
Timp de supraviețuire, 114

Teorie, 1–6, 17, 74, 75
Teste de dezechilibru de transmisie (TDT), 186–189
t statistică, 47–48, 50
valori critice ale, 225
testul t
pentru diferenta intre doua înseamnă, 57–58
pereche egală, 60–61
performant, 58–60
utilizarea, 61–63

Studii pe gemeni, 174–176
ANOVA cu doi factori, 68 exemplu de, 70–71
Test cu două cozi, 58, 148, 157, 159, 225 test cu o singură coadă versus , 147–148
Eroare de tip I, 13–15, 30, 62, 154, 156, 157, 169, 186, 195, 196
consecințele, 15–16
Eroare de tip II, 13–16, 57, 61, 67, 156, 157
consecințele, 15–16

în formă de U, 116, 129–132

Valabilitate, 164, 166–167, 196
Variabilitate, 6–7, 35, 49, 57, 58, 63, 64, 81, 150, 164, 167, 168, 170, 176, 237
Variabile
categoric, 10, 57, 175 confuz, 107–108, 114 continuu, 10, 27, 34, 52, 57, 63, 72, 79, 81, 116, 140, 175
dependent, 76, 79, 93, 100, 107, 111, 114, 115, 119, 121, 125, 129, 170
discret, 10, 27, 34, 57, 81
independent, 68, 69, 76, 79, 80, 92, 100, 107, 109, 110, 115, 118, 119, 170
interacțiunea dintre doi, 69, 125
multiplicitate de, 169
selectând, pentru multivariate modele, 123–125
Varianta
analiza (vezi Analiza varianță)
între grupuri, 64
în cadrul grupurilor, 64
Diagramele Venn, 20, 21

Scanarea întregului genom, 189, 201
Sume de rang Wilcoxon pentru perechi test, 61
Pătratul mediu al grupurilor, 66
Varianta în interiorul grupurilor, 64

Corecția și calculul lui Yates , 30

Scorul Z, 44–47
valori critice de, 56, 225
Testul Z pentru compararea a două proporții, 55–56